

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 13:22:14 ; Search time 34 Seconds
(without alignments)
238.386 Million cell updates/sec

Title: US-10-025-598-2
Perfect score: 190
Sequence: 1 MDWLKARVEQELQALEARGTDSNAELRAMEAKLKAEIQK 39

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 789580 seqs, 207824079 residues

Total number of hits satisfying chosen parameters: 789580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	190	100.0	39	12 US-10-025-598-2	Sequence 2, Appli
2	190	100.0	39	12 US-10-326-441-2	Sequence 2, Appli
3	85	44.7	17	12 US-10-210-023-111	Sequence 111, App
4	60.5	31.8	1549	12 US-10-369-493-5702	Sequence 5702, Ap
5	59.5	31.3	49	15 US-10-023-066A-30	Sequence 30, Appl
6	59.5	31.3	49	15 US-10-023-066A-32	Sequence 32, Appl
7	59.5	31.3	49	15 US-10-023-066A-54	Sequence 54, Appl
8	59.5	31.3	56	15 US-10-023-066A-79	Sequence 79, Appl
9	59.5	31.3	77	15 US-10-023-066A-75	Sequence 75, Appl
10	58.5	30.8	263	10 US-09-925-300-1468	Sequence 1468, Ap
11	58	30.5	30	12 US-09-911-261A-19	Sequence 19, Appl
12	58	30.5	30	12 US-10-173-551-39	Sequence 39, Appl
13	58	30.5	30	15 US-10-057-408-19	Sequence 19, Appl
14	58	30.5	45	9 US-09-766-378A-33	Sequence 33, Appl
15	57.5	30.3	42	15 US-10-023-066A-34	Sequence 34, Appl

16	57.5	30.3	44	11 US-09-491-614-31	Sequence 31, Appl
17	57.5	30.3	56	15 US-10-023-066A-77	Sequence 77, Appl
18	57	30.0	341	12 US-10-192-381-32	Sequence 32, Appl
19	56	29.5	361	9 US-09-841-132-299	Sequence 299, App
20	56	29.5	647	9 US-09-841-132-436	Sequence 436, App
21	55.5	29.2	73	9 US-09-864-761-33395	Sequence 33395, A
22	55.5	29.2	2099	15 US-10-128-714-3290	Sequence 3290, Ap
23	55.5	29.2	2405	15 US-10-128-714-8290	Sequence 8290, Ap
24	55	28.9	28	15 US-10-023-066A-70	Sequence 70, Appl
25	55	28.9	31	10 US-09-756-983-19	Sequence 19, Appl
26	55	28.9	194	12 US-10-211-689-30	Sequence 30, Appl
27	55	28.9	364	10 US-09-756-983-22	Sequence 22, Appl
28	55	28.9	445	15 US-10-156-761-14752	Sequence 14752, A
29	55	28.9	545	12 US-09-849-602-25	Sequence 25, Appl
30	55	28.9	572	12 US-09-738-630-105	Sequence 105, App
31	54.5	28.7	222	9 US-09-840-787-41	Sequence 41, Appl
32	54.5	28.7	663	12 US-10-369-493-19140	Sequence 19140, A
33	54.5	28.7	854	11 US-09-770-107-2	Sequence 2, Appli
34	54	28.4	74	9 US-09-864-761-45412	Sequence 45412, A
35	54	28.4	454	12 US-10-369-493-23078	Sequence 23078, A
36	54	28.4	660	12 US-09-841-260-139	Sequence 139, App
37	54	28.4	660	14 US-10-007-693-139	Sequence 139, App
38	54	28.4	1091	12 US-10-369-493-6328	Sequence 6328, Ap
39	54	28.4	2383	15 US-10-082-830-260	Sequence 260, App
40	54	28.4	5795	9 US-09-815-242-12610	Sequence 12610, A
41	53.5	28.2	1118	12 US-10-369-493-18845	Sequence 18845, A
42	53.5	28.2	2288	12 US-10-369-493-6774	Sequence 6774, Ap
43	53.5	28.2	3500	15 US-10-153-219-2	Sequence 2, Appli
44	53.5	28.2	3537	15 US-10-153-219-15	Sequence 15, Appl
45	53	27.9	151	9 US-09-840-787-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-10-025-598-2
; Sequence 2, Application US/10025598
; Publication No. US20030147852A1
; GENERAL INFORMATION:
; APPLICANT: Schaaak, Diane L.
; TITLE OF INVENTION: TOXIN-PHAGE BACTERIOCID ANTIBIOTIC AND USES THEREOF
; FILE REFERENCE: 13688-002001
; CURRENT APPLICATION NUMBER: US/10/025,598
; CURRENT FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetically generated peptide
US-10-025-598-2

Query Match 100.0%; Score 190; DB 12; Length 39;
Best Local Similarity 100.0%; Pred. No. 6.1e-18;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MDWLKARVEQELQALEARGTDSNAELRAMEAKLKAEIQK 39
Db 1 MDWLKARVEQELQALEARGTDSNAELRAMEAKLKAEIQK 39

RESULT 2
US-10-326-441-2
; Sequence 2, Application US/10326441
; Publication No. US20030175246A1
; GENERAL INFORMATION:
; APPLICANT: Schaaak, Diane L.
; TITLE OF INVENTION: TOXIN-PHAGE BACTERIOCID ANTIBIOTIC AND
; FILE REFERENCE: 00246-269001

Search Notes
10/025,598
2/21/04

Db 12 LKA-MEEKLKAMEEKLKAMEEKLKAMEEK 46

RESULT 6

US-10-023-066A-32

; Sequence 32, Application US/10023066A

; Publication No. US20030056242A1

; GENERAL INFORMATION:

; APPLICANT: E. I. DU PONT DE NEMOURS AND COMPANY

; TITLE OF INVENTION: CHIMERIC GENES AND METHODS FOR INCREASING THE LYSINE AND THREONINE CONTENT OF THE SEEDS OF PLANTS

; NUMBER OF SEQUENCES: 107

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY

; STREET: 1007 MARKET STREET

; CITY: WILMINGTON

; STATE: DELAWARE

; COUNTRY: U.S.A.

; ZIP: 19898

; COMPUTER READABLE FORM:

; MEDIUM TYPE: FLOPPY DISK

; COMPUTER: IBM PC COMPATIBLE

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: MICROSOFT WORD VERSION 2.0C

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/023,066A

; FILING DATE: 29-Apr-2002

; CLASSIFICATION: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: BARBARA C. SIEGELL

; REGISTRATION NUMBER: 30,684

; REFERENCE/DOCKET NUMBER: BB-1037-C

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 302-992-4931

; TELEFAX: 302-773-0164

; TELEX: 835420

; INFORMATION FOR SEQ ID NO: 32:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 49 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 32:

US-10-023-066A-32

Query Match 31.3%; Score 59.5; DB 15; Length 49;

Best Local Similarity 44.4%; Pred. No. 1;

Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

Qy 4 LKARVEQELQALEARGTDSNAELRAMEAKLKAEIQK 39

Db 12 LKA-MEEKLKAMEEKLKAMEEKLKAMEEK 46

RESULT 7

US-10-023-066A-54

; Sequence 54, Application US/10023066A

; Publication No. US20030056242A1

; GENERAL INFORMATION:

; APPLICANT: E. I. DU PONT DE NEMOURS AND COMPANY

; TITLE OF INVENTION: CHIMERIC GENES AND METHODS FOR INCREASING THE LYSINE AND THREONINE CONTENT OF THE SEEDS OF PLANTS

; NUMBER OF SEQUENCES: 107

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY

STREET: 1007 MARKET STREET

CITY: WILMINGTON

STATE: DELAWARE

COUNTRY: U.S.A.

ZIP: 19898

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: MICROSOFT WORD VERSION 2.0C

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/023,066A

FILING DATE: 29-Apr-2002

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: BARBARA C. SIEGELL

REGISTRATION NUMBER: 30,684

REFERENCE/DOCKET NUMBER: BB-1037-C

TELECOMMUNICATION INFORMATION:

TELEPHONE: 302-992-4931

TELEFAX: 302-773-0164

TELEX: 835420

INFORMATION FOR SEQ ID NO: 54:

SEQUENCE CHARACTERISTICS:

LENGTH: 49 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 54:

US-10-023-066A-54

Query Match 31.3%; Score 59.5; DB 15; Length 49;

Best Local Similarity 44.4%; Pred. No. 1;

Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

Qy 4 LKARVEQELQALEARGTDSNAELRAMEAKLKAEIQK 39

Db 5 LKA-MEEKLKAMEEKLKAMEEKLKAMEEK 39

RESULT 8

US-10-023-066A-79

; Sequence 79, Application US/10023066A

; Publication No. US20030056242A1

; GENERAL INFORMATION:

; APPLICANT: E. I. DU PONT DE NEMOURS AND COMPANY

; TITLE OF INVENTION: CHIMERIC GENES AND METHODS FOR INCREASING THE LYSINE AND THREONINE CONTENT OF THE SEEDS OF PLANTS

; NUMBER OF SEQUENCES: 107

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY

; STREET: 1007 MARKET STREET

; CITY: WILMINGTON

; STATE: DELAWARE

; COUNTRY: U.S.A.

; ZIP: 19898

; COMPUTER READABLE FORM:

; MEDIUM TYPE: FLOPPY DISK

; COMPUTER: IBM PC COMPATIBLE

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: MICROSOFT WORD VERSION 2.0C

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/023,066A

; FILING DATE: 29-Apr-2002

; CLASSIFICATION: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: BARBARA C. SIEGELL

; REGISTRATION NUMBER: 30,684

; REFERENCE/DOCKET NUMBER: BB-1037-C

```
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4931
; TELEFAX: 302-773-0164
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 56 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-10-023-066A-79
Query Match 31.3%; Score 59.5; DB 15; Length 56;
Best Local Similarity 44.4%; Pred. No. 1.2;
Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

Qy 4 LKARVEQLQALEARGTDSNAELRAMEAKLKAEIQK 39
Db 5 LKA-MEEKLKAMEEKLKAMEEKLKAMEEKLKAMEEK 39

RESULT 9
US-10-023-066A-75
; Sequence 75, Application US/10023066A
; Publication No. US20030056242A1
; GENERAL INFORMATION:
; APPLICANT: E. I. DU PONT DE NEMOURS AND COMPANY
; TITLE OF INVENTION: CHIMERIC GENES AND METHODS FOR INCREASING THE LYSINE AND THREONINE CONTENT OF THE SEEDS OF PLANTS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD VERSION 2.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/023,066A
; FILING DATE: 29-Apr-2002
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: BARBARA C. SIEGELL
; REGISTRATION NUMBER: 30,684
; REFERENCE/DOCKET NUMBER: BB-1037-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4931
; TELEFAX: 302-773-0164
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 77 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 75:
US-10-023-066A-75
Query Match 31.3%; Score 59.5; DB 15; Length 77;
Best Local Similarity 44.4%; Pred. No. 1.7;
Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

Qy 4 LKARVEQLQALEARGTDSNAELRAMEAKLKAEIQK 39
Db 5 LKA-MEEKLKAMEEKLKAMEEKLKAMEEKLKAMEEK 39

;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4931
; TELEFAX: 302-773-0164
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 56 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-10-023-066A-79
Query Match 31.3%; Score 59.5; DB 15; Length 56;
Best Local Similarity 44.4%; Pred. No. 1.2;
Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

Qy 4 LKARVEQLQALEARGTDSNAELRAMEAKLKAEIQK 39
Db 5 LKA-MEEKLKAMEEKLKAMEEKLKAMEEKLKAMEEK 39

RESULT 10
US-09-925-300-1468
; Sequence 1468, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1468
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (6)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-300-1468
Query Match 30.8%; Score 58.5; DB 10; Length 263;
Best Local Similarity 36.8%; Pred. No. 9.4;
Matches 14; Conservative 11; Mismatches 12; Indels 1; Gaps 1;

Qy 2 DWLKARVEQLQALEARGT-DSNAELRAMEAKLKAEIQ 38
Db 80 EFLEKKIEQELTAARKHGKTKNRAALQALKRKRYEKQ 117

RESULT 11
US-09-911-261A-19
; Sequence 19, Application US/09911261A
; Publication No. US20030134350A1
; GENERAL INFORMATION:
; APPLICANT: Sera, Takashi
; TITLE OF INVENTION: Zinc Finger Domain Recognition Code and Uses Thereof
; FILE REFERENCE: 109845.135
; CURRENT APPLICATION NUMBER: US/09/911,261A
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: US 60/220,060
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 19
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Acid dimerization peptide.
US-09-911-261A-19
Query Match 30.5%; Score 58; DB 12; Length 30;
Best Local Similarity 46.4%; Pred. No. 0.9;
Matches 13; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

Qy 6 ARVEQLQALEARGTDSNAELRAMEAKL 33
Db 1 AQLEKELQALEKENAQLEWELQALEKEL 28

RESULT 12
US-10-173-551-39
```

; Sequence 39, Application US/10173551
; Publication No. US20030232387A1
; GENERAL INFORMATION:
; APPLICANT: Lu, Chafen
; TITLE OF INVENTION: Antibodies that bind alphaE Integrin
; FILE REFERENCE: 1855.2025-000
; CURRENT APPLICATION NUMBER: US/10/173,551
; CURRENT FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: acidic peptide
US-10-173-551-39

Query Match 30.5%; Score 58; DB 12; Length 30;
Best Local Similarity 46.4%; Pred. No. 0.9;
Matches 13; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 6 ARVEQLQALEARGTDSNAELRAMEAKL 33
|:::||||| ||:::|
Db 1 AQLEKELQALEKENAQLEWELQALEKEL 28

RESULT 13
US-10-057-408-19
; Sequence 19, Application US/10057408
; Publication No. US20030082561A1
; GENERAL INFORMATION:
; APPLICANT: Sera, Takashi
; TITLE OF INVENTION: Zinc Finger Domain Recognition Code and Uses Thereof
; FILE REFERENCE: 109845.135
; CURRENT APPLICATION NUMBER: US/10/057,408
; CURRENT FILING DATE: 2002-01-23
; PRIOR APPLICATION NUMBER: US 60/220,060
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 19
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Acid dimerization peptide.
US-10-057-408-19

Query Match 30.5%; Score 58; DB 15; Length 30;
Best Local Similarity 46.4%; Pred. No. 0.9;
Matches 13; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 6 ARVEQLQALEARGTDSNAELRAMEAKL 33
|:::||||| ||:::|
Db 1 AQLEKELQALEKENAQLEWELQALEKEL 28

RESULT 14
US-09-766-378A-33
; Sequence 33, Application US/09766378A
; Patent No. US20020091079A1
; GENERAL INFORMATION:
; APPLICANT: Rhode, Peter R.
; Acevedo, Jorge
; Burkhardt, Martin
; Jiao, Jin-an
; Wong, Hing C.
; TITLE OF INVENTION: SOLUBLE MHC COMPLEXES AND METHODS OF USE THEREOF
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dike, Bronstein, Roberts & Cushman, LLP

; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: usa
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/766,378A
; FILING DATE: 19-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/960,190
; FILING DATE: 29-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Corless, Peter F
; REGISTRATION NUMBER: 33,860
; REFERENCE/DOCKET NUMBER: 48002-DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 33:
US-09-766-378A-33

Query Match 30.5%; Score 58; DB 9; Length 45;
Best Local Similarity 46.4%; Pred. No. 1.4;
Matches 13; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 6 ARVEQLQALEARGTDSNAELRAMEAKL 33
|:::||||| ||:::|
Db 16 AQLEKELQALEKENAQLEWELQALEKEL 43

RESULT 15
US-10-023-066A-34
; Sequence 34, Application US/10023066A
; Publication No. US20030056242A1
; GENERAL INFORMATION:
; APPLICANT: E. I. DU PONT DE NEMOURS AND COMPANY
; TITLE OF INVENTION: CHIMERIC GENES AND METHODS FOR INCREASING THE LYSINE AND THREONINE CONTENT OF THE SEEDS OF PLANTS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD VERSION 2.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/023,066A
; FILING DATE: 29-Apr-2002
; CLASSIFICATION: <Unknown>

```

;
; ATTORNEY/AGENT INFORMATION:
; NAME: BARBARA C. SIEGELL
; REGISTRATION NUMBER: 30,684
; REFERENCE/DOCKET NUMBER: BB-1037-C
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4931
; TELEFAX: 302-773-0164
; TELEX: 835420
;
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-10-023-066A-34

Query Match 30.3%; Score 57.5; DB 15; Length 42;
Best Local Similarity 41.7%; Pred. No. 1.5;
Matches 15; Conservative 10; Mismatches 10; Indels 1; Gaps 1;

Qy 4 LKARVEQLQALEARGTDSNAELRAMEAKLKAEIQK 39
Db 5 MKA-MEEKLKAMEEKLKAMEEKLKAMEEKLKAMEEK 39

```

Search completed: January 30, 2004, 13:27:45
Job time : 34 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 13:20:14 ; Search time 22 Seconds
(without alignments)
75.006 Million cell updates/sec

Title: US-10-025-598-2
Perfect score: 190
Sequence: 1 MDWLKARVEQLQALEARGTDSNAELRAMEAKLKAEIQK 39

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	59.5	31.3	49	1 US-08-182-175A-33	Sequence 33, Appl
2	59.5	31.3	49	1 US-08-182-175A-35	Sequence 35, Appl
3	59.5	31.3	49	1 US-08-182-175A-45	Sequence 45, Appl
4	59.5	31.3	49	1 US-08-474-633A-30	Sequence 30, Appl
5	59.5	31.3	49	1 US-08-474-633A-32	Sequence 32, Appl
6	59.5	31.3	49	1 US-08-474-633A-54	Sequence 54, Appl
7	59.5	31.3	49	4 US-08-823-771-30	Sequence 30, Appl
8	59.5	31.3	49	4 US-08-823-771-32	Sequence 32, Appl
9	59.5	31.3	49	4 US-08-823-771-54	Sequence 54, Appl
10	59.5	31.3	49	5 PCT-US92-06412-33	Sequence 33, Appl
11	59.5	31.3	49	5 PCT-US92-06412-35	Sequence 35, Appl
12	59.5	31.3	49	5 PCT-US92-06412-45	Sequence 45, Appl
13	59.5	31.3	56	1 US-08-182-175A-91	Sequence 91, Appl
14	59.5	31.3	56	1 US-08-182-175A-112	Sequence 112, Appl
15	59.5	31.3	56	1 US-08-474-633A-79	Sequence 79, Appl
16	59.5	31.3	56	4 US-08-823-771-79	Sequence 79, Appl
17	59.5	31.3	56	5 PCT-US92-06412-91	Sequence 91, Appl
18	59.5	31.3	56	5 PCT-US92-06412-112	Sequence 112, Appl
19	59.5	31.3	77	1 US-08-182-175A-57	Sequence 57, Appl
20	59.5	31.3	77	1 US-08-474-633A-75	Sequence 75, Appl
21	59.5	31.3	77	4 US-08-823-771-75	Sequence 75, Appl
22	59.5	31.3	77	5 PCT-US92-06412-57	Sequence 57, Appl
23	58	30.5	45	3 US-08-960-190A-33	Sequence 33, Appl
24	57.5	30.3	42	1 US-08-182-175A-37	Sequence 37, Appl
25	57.5	30.3	42	1 US-08-474-633A-34	Sequence 34, Appl
26	57.5	30.3	42	4 US-08-823-771-34	Sequence 34, Appl
27	57.5	30.3	42	5 PCT-US92-06412-37	Sequence 37, Appl

28	57.5	30.3	56	1 US-08-182-175A-59	Sequence 59, Appl
29	57.5	30.3	56	1 US-08-474-633A-77	Sequence 77, Appl
30	57.5	30.3	56	4 US-08-823-771-77	Sequence 77, Appl
31	57.5	30.3	56	5 PCT-US92-06412-59	Sequence 59, Appl
32	57	30.0	220	4 US-09-328-352-5166	Sequence 5166, Ap
33	56	29.5	361	4 US-09-556-877-299	Sequence 299, App
34	56	29.5	361	4 US-09-620-412C-299	Sequence 299, App
35	56	29.5	361	4 US-09-598-419-299	Sequence 299, App
36	55.5	29.2	167	4 US-09-107-532A-6197	Sequence 6197, Ap
37	55.5	29.2	284	3 US-09-053-197A-2	Sequence 2, Appli
38	55.5	29.2	284	4 US-09-085-761A-2	Sequence 2, Appli
39	55.5	29.2	584	1 US-08-179-738-7	Sequence 7, Appli
40	55.5	29.2	584	2 US-08-628-145-7	Sequence 7, Appli
41	55.5	29.2	591	1 US-08-179-738-5	Sequence 5, Appli
42	55.5	29.2	591	1 US-08-179-738-10	Sequence 10, Appl
43	55.5	29.2	591	2 US-08-628-145-5	Sequence 5, Appli
44	55.5	29.2	591	2 US-08-628-145-10	Sequence 10, Appl
45	55.5	29.2	595	1 US-08-171-718-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-08-182-175A-33
; Sequence 33, Application US/08182175A
; Patent No. 5559223
; GENERAL INFORMATION:
; APPLICANT: Saverio Carl Falco
; APPLICANT: Sharon J. Keeler
; APPLICANT: Janet A. Rice
; TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing P
; NUMBER OF SEQUENCES: 113
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. du Pont de Nemours and Company
; STREET: 1007 Market Street
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: USA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh System, 6.0
; SOFTWARE: Microsoft Word, 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/182,175A
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/743,006
; FILING DATE: 9 August 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Linda Axamethy Floyd
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: BB-1031
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 992-4929
; TELEFAX: (302) 892-7949
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 49 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-182-175A-33

Query Match 31.3%; Score 59.5; DB 1; Length 49;
Best Local Similarity 44.4%; Pred. No. 0.41;
Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

QY 4 LKARVEQLQALEARGTDSNAELRAMEAKLKAEIQK 39

10/025,598
Search Notes
2/2/04

STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: U.S.A.
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD VERSION 2.0C
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/823,771
FILING DATE: 24-Mar-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/474,633
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BARBARA C. SIEGELL
REGISTRATION NUMBER: 30,684
REFERENCE/DOCKET NUMBER: BB-1037-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-992-4931
TELEFAX: 302-773-0164
TELEX: 835420
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 49 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 30:
US-08-823-771-30
Query Match 31.3%; Score 59.5; DB 4; Length 49;
Best Local Similarity 44.4%; Pred. No. 0.41;
Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;
QY 4 LKARVEQLQALEARGTDSNAELRAMEAKLKAEIQK 39
||| :|:|:|:| :|:|:|:|:|:| :|
Db 12 LKA-MEEKLKAMEEKLKAMEEKLKAMEEKLKAMEEK 46
RESULT 8
US-08-823-771-32
Sequence 32, Application US/08823771
Patent No. 6459019
GENERAL INFORMATION:
APPLICANT: E. I. DU PONT DE NEMOURS AND COMPANY
TITLE OF INVENTION: CHIMERIC GENES AND METHODS FOR INCREASING INCREASING THE LYSINE AND THREONINE CONTENT
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: U.S.A.
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD VERSION 2.0C
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/823,771
FILING DATE: 24-Mar-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/474,633
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BARBARA C. SIEGELL
REGISTRATION NUMBER: 30,684
REFERENCE/DOCKET NUMBER: BB-1037-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-992-4931
TELEFAX: 302-773-0164
TELEX: 835420
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 49 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 32:
US-08-823-771-32
Query Match 31.3%; Score 59.5; DB 4; Length 49;
Best Local Similarity 44.4%; Pred. No. 0.41;
Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;
QY 4 LKARVEQLQALEARGTDSNAELRAMEAKLKAEIQK 39
||| :|:|:|:| :|:|:|:|:|:| :|
Db 12 LKA-MEEKLKAMEEKLKAMEEKLKAMEEKLKAMEEK 46
RESULT 9
US-08-823-771-54
Sequence 54, Application US/08823771
Patent No. 6459019
GENERAL INFORMATION:
APPLICANT: E. I. DU PONT DE NEMOURS AND COMPANY
TITLE OF INVENTION: CHIMERIC GENES AND METHODS FOR INCREASING INCREASING THE LYSINE AND THREONINE CONTENT
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: U.S.A.
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD VERSION 2.0C
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/823,771
FILING DATE: 24-Mar-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/474,633
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BARBARA C. SIEGELL
REGISTRATION NUMBER: 30,684
REFERENCE/DOCKET NUMBER: BB-1037-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-992-4931
TELEFAX: 302-773-0164
TELEX: 835420
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 49 amino acids
TYPE: amino acid
TOPOLOGY: linear

```
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 54:
US-08-823-771-54

Query Match          31.3%; Score 59.5; DB 4; Length 49;
Best Local Similarity 44.4%; Pred. No. 0.41;
Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

Qy 4 LKARVEQLQALEARGTDSNAELRAMEAKLKAEIQK 39
    ||| :|::|:| : :|::| | | | :|
Db 5 LKA-MEEKLKAMEEKLKAMEEKLKAMEEKLKAMEEK 39

RESULT 10
PCT-US92-06412-33
; Sequence 33, Application PC/TUS9206412
; GENERAL INFORMATION:
; APPLICANT: Saverio Carl Falco
; APPLICANT: Sharon J. Keeler
; APPLICANT: Janet A. Rice
; TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing B
; NUMBER OF SEQUENCES: 113
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E.I. du Pont de Nemours and Company
; STREET: 1007 Market Street
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: USA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh System, 6.0
; SOFTWARE: Microsoft Word, 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06412
; FILING DATE: 19920807
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/743,006
; FILING DATE: 9 August 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Linda Axamethy Floyd
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: BB-1031
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 992-4929
; TELEFAX: (302) 892-7949
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 49 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US92-06412-33

Query Match          31.3%; Score 59.5; DB 5; Length 49;
Best Local Similarity 44.4%; Pred. No. 0.41;
Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

Qy 4 LKARVEQLQALEARGTDSNAELRAMEAKLKAEIQK 39
    ||| :|::|:| : :|::| | | | :|
Db 12 LKA-MEEKLKAMEEKLKAMEEKLKAMEEKLKAMEEK 46

RESULT 11
PCT-US92-06412-35
; Sequence 35, Application PC/TUS9206412
; GENERAL INFORMATION:
; APPLICANT: Saverio Carl Falco
; APPLICANT: Sharon J. Keeler
; APPLICANT: Janet A. Rice
```

```
; TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing P
; NUMBER OF SEQUENCES: 113
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E.I. du Pont de Nemours and Company
; STREET: 1007 Market Street
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: USA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh System, 6.0
; SOFTWARE: Microsoft Word, 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06412
; FILING DATE: 19920807
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/743,006
; FILING DATE: 9 August 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Linda Axamethy Floyd
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: BB-1031
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 992-4929
; TELEFAX: (302) 892-7949
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 49 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US92-06412-35

Query Match          31.3%; Score 59.5; DB 5; Length 49;
Best Local Similarity 44.4%; Pred. No. 0.41;
Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

Qy 4 LKARVEQLQALEARGTDSNAELRAMEAKLKAEIQK 39
    ||| :|::|:| : :|::| | | | :|
Db 12 LKA-MEEKLKAMEEKLKAMEEKLKAMEEKLKAMEEK 46

RESULT 12
PCT-US92-06412-45
; Sequence 45, Application PC/TUS9206412
; GENERAL INFORMATION:
; APPLICANT: Saverio Carl Falco
; APPLICANT: Sharon J. Keeler
; APPLICANT: Janet A. Rice
; TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing P
; NUMBER OF SEQUENCES: 113
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E.I. du Pont de Nemours and Company
; STREET: 1007 Market Street
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: USA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh System, 6.0
; SOFTWARE: Microsoft Word, 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06412
; FILING DATE: 19920807
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/743,006
```

```

; FILING DATE: 9 August 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Linda Axamethy Floyd
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: BB-1031
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 992-4929
; TELEFAX: (302) 892-7949
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 49 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US92-06412-45

Query Match 31.3%; Score 59.5; DB 5; Length 49;
Best Local Similarity 44.4%; Pred. No. 0.41;
Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

QY 4 LKARVEQELQALEARGTDSNAELRAMEAKLKAEIQK 39
Db 5 LKA-MEEKLKAMEEKLKAMEEKLKAMEEKLKAMEEK 39

RESULT 13
US-08-182-175A-91
; Sequence 91, Application US/08182175A
; Patent No. 5559223
; GENERAL INFORMATION:
; APPLICANT: Saverio Carl Falco
; APPLICANT: Sharon J. Keeler
; APPLICANT: Janet A. Rice
; TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing P
; NUMBER OF SEQUENCES: 113
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E.I. du Pont de Nemours and Company
; STREET: 1007 Market Street
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: USA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh System, 6.0
; SOFTWARE: Microsoft Word, 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/182,175A
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/743,006
; FILING DATE: 9 August 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Linda Axamethy Floyd
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: BB-1031
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 992-4929
; TELEFAX: (302) 892-7949
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 56 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-182-175A-91

Query Match 31.3%; Score 59.5; DB 1; Length 56;
Best Local Similarity 44.4%; Pred. No. 0.48;
Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

QY 4 LKARVEQELQALEARGTDSNAELRAMEAKLKAEIQK 39
Db 5 LKA-MEEKLKAMEEKLKAMEEKLKAMEEKLKAMEEK 39

RESULT 14
US-08-182-175A-112
; Sequence 112, Application US/08182175A
; Patent No. 5559223
; GENERAL INFORMATION:
; APPLICANT: Saverio Carl Falco
; APPLICANT: Sharon J. Keeler
; APPLICANT: Janet A. Rice
; TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing P
; NUMBER OF SEQUENCES: 113
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E.I. du Pont de Nemours and Company
; STREET: 1007 Market Street
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: USA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh System, 6.0
; SOFTWARE: Microsoft Word, 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/182,175A
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/743,006
; FILING DATE: 9 August 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Linda Axamethy Floyd
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: BB-1031
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 992-4929
; TELEFAX: (302) 892-7949
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 112:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 56 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..56
; OTHER INFORMATION: /label= name
; OTHER INFORMATION: /note= "SSP-3- 5(A/E)"
; US-08-182-175A-112

Query Match 31.3%; Score 59.5; DB 1; Length 56;
Best Local Similarity 44.4%; Pred. No. 0.48;
Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

QY 4 LKARVEQELQALEARGTDSNAELRAMEAKLKAEIQK 39
Db 5 LKA-MEEKLKAMEEKLKAMEEKLKAMEEKLKAMEEK 39

RESULT 15
US-08-474-633A-79
; Sequence 79, Application US/08474633A
; Patent No. 5773691
; GENERAL INFORMATION:
; APPLICANT: E. I. DU PONT DE NEMOURS AND
```

```

Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

QY 4 LKARVEQELQALEARGTDSNAELRAMEAKLKAEIQK 39
Db 5 LKA-MEEKLKAMEEKLKAMEEKLKAMEEKLKAMEEK 39

RESULT 14
US-08-182-175A-112
; Sequence 112, Application US/08182175A
; Patent No. 5559223
; GENERAL INFORMATION:
; APPLICANT: Saverio Carl Falco
; APPLICANT: Sharon J. Keeler
; APPLICANT: Janet A. Rice
; TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing P
; NUMBER OF SEQUENCES: 113
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E.I. du Pont de Nemours and Company
; STREET: 1007 Market Street
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: USA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh System, 6.0
; SOFTWARE: Microsoft Word, 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/182,175A
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/743,006
; FILING DATE: 9 August 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Linda Axamethy Floyd
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: BB-1031
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 992-4929
; TELEFAX: (302) 892-7949
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 112:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 56 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..56
; OTHER INFORMATION: /label= name
; OTHER INFORMATION: /note= "SSP-3- 5(A/E)"
; US-08-182-175A-112

Query Match 31.3%; Score 59.5; DB 1; Length 56;
Best Local Similarity 44.4%; Pred. No. 0.48;
Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

QY 4 LKARVEQELQALEARGTDSNAELRAMEAKLKAEIQK 39
Db 5 LKA-MEEKLKAMEEKLKAMEEKLKAMEEKLKAMEEK 39

RESULT 15
US-08-474-633A-79
; Sequence 79, Application US/08474633A
; Patent No. 5773691
; GENERAL INFORMATION:
; APPLICANT: E. I. DU PONT DE NEMOURS AND
```

Search completed: January 30, 2004, 13:23:19
Job time : 23 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 13:18:38 ; Search time 35 Seconds
(without alignments)
287.544 Million cell updates/sec

Title: US-10-025-598-2
Perfect score: 190
Sequence: 1 MDWLKARVEQELQALEARGTDSNAELRAMEAKLKAEIQK 39

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL_23.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phage.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvirus.*
 - 16: sp_bacteriap.*
 - 17: sp_archheap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62	32.6	210	16	Q8XXF0 ralstonia s
2	60.5	31.8	1549	5	Q20060 caenorhabdi
3	60	31.6	680	11	Q8R370 mus musculu
4	59.5	31.3	594	11	Q8CHP2 mus musculu
5	59	31.1	352	10	Q94GZ7 oryza sativ
6	58.5	30.8	705	10	Q9SU64 arabidopsis
7	58.5	30.8	737	16	Q9KTG5 vibrio chol
8	58.5	30.8	788	11	Q8CF87 mus musculu
9	58.5	30.8	851	11	Q8CF88 mus musculu
10	58.5	30.8	909	10	Q9SU09 arabidopsis
11	58	30.5	358	11	Q9Z2X5 rattus norv
12	58	30.5	1941	5	Q26079 placoepten
13	58	30.5	1950	5	Q26080 placoepten
14	58	30.5	2142	10	Q9FWR3 arabidopsis
15	57.5	30.3	122	17	Q97Z14 sulfolobus
16	57.5	30.3	153	10	Q9M2J3 arabidopsis

17	57.5	30.3	259	16	Q9RZ09	deinococcus
18	57.5	30.3	661	12	Q9PYR3	xestia c-ni
19	57.5	30.3	998	16	Q8D0Z4	yersinia pe
20	57.5	30.3	1052	2	Q9F291	yersinia pe
21	57	30.0	306	16	Q8PPJ5	xanthomonas
22	57	30.0	356	11	Q99JP6	mus musculu
23	57	30.0	358	11	Q9Z215	mus musculu
24	57	30.0	1052	11	Q9ERA1	microtus ar
25	57	30.0	1243	11	Q9ERA5	microtus ar
26	56.5	29.7	174	12	Q86999	soil-borne
27	56.5	29.7	334	16	Q9RJS9	streptomyce
28	56.5	29.7	836	11	Q8VDR2	mus musculu
29	56	29.5	211	16	Q8G4L3	bifidobacte
30	56	29.5	227	16	Q8EG07	shewanella
31	56	29.5	353	16	Q9A776	caulobacter
32	56	29.5	504	16	Q8PAS9	xanthomonas
33	56	29.5	647	16	Q84627	chlamydia t
34	56	29.5	665	11	Q91WF2	mus musculu
35	56	29.5	665	11	Q9DC21	mus musculu
36	56	29.5	842	3	Q96WP6	lentinula e
37	56	29.5	1078	10	Q8LI05	oryza sativ
38	56	29.5	1945	5	Q9BLD0	patinopecte
39	55.5	29.2	244	3	Q74734	schizosacch
40	55.5	29.2	279	4	Q96T30	homo sapien
41	55.5	29.2	499	4	Q9BTW3	homo sapien
42	55.5	29.2	507	4	Q96T31	homo sapien
43	55.5	29.2	510	11	Q8KIY6	mus musculu
44	55.5	29.2	546	11	Q9D3K3	mus musculu
45	55.5	29.2	548	4	Q96T33	homo sapien

ALIGNMENTS

RESULT 1

Q8XXF0	ID	Q8XXF0	PRELIMINARY;	PRT;	210 AA.
AC	Q8XXF0;				
DT	01-MAR-2002 (TREMBlrel. 20, Created)				
DT	01-MAR-2002 (TREMBlrel. 20, Last sequence update)				
DT	01-OCT-2002 (TREMBlrel. 22, Last annotation update)				
DE	Probable signal peptide protein.				
GN	RSC2164 OR RS01436.				
OS	Ralstonia solanacearum (Pseudomonas solanacearum).				
OC	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;				
OC	Ralstoniaceae; Ralstonia.				
OX	NCBI_TaxID=305;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=GMI1000;				
RX	MEDLINE=21681879; PubMed=11823852;				
RA	Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,				
RA	Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,				
RA	Chandler M., Choisine N., Claudel-Renard C., Cunnac S., Demange N.,				
RA	Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,				
RA	Siguiet P., Thebault P., Whalen M., Wincker P., Levy M.,				
RA	Weissenbach J., Boucher C.A.;				
RT	"Genome sequence of the plant pathogen Ralstonia solanacearum.";				
RL	Nature 415:497-502(2002).				
DR	EMBL; AL646068; CAD15871.1; --				
KW	Complete proteome.				
SQ	SEQUENCE 210 AA; 23230 MW; 482BB9EF959BA875 CRC64;				

Query Match 32.6%; Score 62; DB 16; Length 210;
Best Local Similarity 37.1%; Pred. No. 13;
Matches 13; Conservative 11; Mismatches 11; Indels 0; Gaps

QY	5	KARVEQELQALEARGTDSNAELRAMEAKLKAEIQK 39
Db	131	KAQLEREQAEEQRAAALSQEVRAQARQOMLEE 165

RESULT 2

10/025,598
Search Notes
2/21/04

AC Q9SU09;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Auxilin-like protein.
 GN T20K18.120 OR AT4G12770.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Peters S.A., van Staveren M., Dirkse W., Stiekema W.,
 RA Bancroft I., Mewes H.W., Mayer K.F.X., Schueller C.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL049640; CAB40994.1; -.
 DR EMBL; AL161534; CAB78319.1; -.
 DR InterPro; IPR001623; DnaJ_N.
 DR InterPro; IPR002965; P rich extensn.
 DR PRINTS; PR01217; PRICHEXTENS.
 DR SMART; SM00271; DnaJ; 1.
 SQ SEQUENCE 909 AA; 100307 MW; 692138C99F32DFFC CRC64;
 Query Match 30.8%; Score 58.5; DB 10; Length 909;
 Best Local Similarity 51.6%; Pred. No. 1.5e+02;
 Matches 16; Conservative 5; Mismatches 9; Indels 1; Gaps 1;
 QY 6 ARVEQELQALEARGTDSNAELRAMEAKLAE 36
 Db 605 AREKAEKAAAEAR-ERANAEEVREKEAKVRAE 634
 RESULT 11
 Q9Z2X5 PRELIMINARY; PRT; 358 AA.
 AC Q9Z2X5;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Vesl-3.
 GN VESL-3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Hippocampus;
 RA Kato A., Ageta H., Sugiyama H.;
 RT "Rattus norvegicus mRNA for Vesl-3, complete cds."
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB020879; BAA35110.2; -.
 DR InterPro; IPR000697; EVH1.
 DR InterPro; IPR001960; WH1.
 DR Pfam; PF00568; WH1; 1.
 DR SMART; SM00461; WH1; 1.
 SQ SEQUENCE 358 AA; 39891 MW; 4896E23D6A63517E CRC64;
 Query Match 30.5%; Score 58; DB 11; Length 358;
 Best Local Similarity 36.7%; Pred. No. 66;
 Matches 11; Conservative 8; Mismatches 11; Indels 0; Gaps 0;
 QY 5 KARVEQLQALEARGTDSNAELRAMEAKLK 34
 : :|:|:| | | : :|:|:| | | :

Db 287 REETQQQVQDLETRNAELEQLRAMESNLE 316
 RESULT 12
 Q26079 PRELIMINARY; PRT; 1941 AA.
 AC Q26079;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Myosin heavy chain.
 OS Placopecten magellanicus (Sea scallop).
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Pectinoidea;
 OC Pectinoidea; Pectinidae; Placopecten.
 OX NCBI_TaxID=6577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skeletal muscle;
 RA Perreault-Micale C., Kalabokis V., Nyitray L., Szent-Gyorgyi A.G.;
 RT "Sequence variations in the surface loop near the nucleotide binding
 RT site modulate the ATP turnover rates of molluscan myosins.";
 RL J. Muscle Res. Cell Motil. 0:0-0(1996).
 DR EMBL; U59294; AAB03660.1; -.
 DR HSSP; P24733; IWDC.
 DR InterPro; IPR000048; IQ region.
 DR InterPro; IPR001609; myosin_head.
 DR InterPro; IPR004009; Myosin_N.
 DR InterPro; IPR002928; Myosin_tail.
 DR InterPro; IPR002017; Spectrin.
 DR Pfam; PF00612; IQ; 2.
 DR Pfam; PF00063; myosin_head; 1.
 DR Pfam; PF02736; Myosin_N; 1.
 DR Pfam; PF01576; Myosin_tail; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SM00015; IQ; 1.
 DR SMART; SM00242; MYSC; 1.
 DR PROSITE; PS50096; IQ; 1.
 SQ SEQUENCE 1941 AA; 223243 MW; E2AD1637FFDB8127 CRC64;
 Query Match 30.5%; Score 58; DB 5; Length 1941;
 Best Local Similarity 37.1%; Pred. No. 3.9e+02;
 Matches 13; Conservative 8; Mismatches 14; Indels 0; Gaps 0;
 QY 1 MDWLKARVEQLQALEARGTDSNAELRAMEAKLKA 35
 Db 1414 LDKAKSRLQCELEDLSIEVDNANVNQMEKKQRA 1448
 RESULT 13
 Q26080 PRELIMINARY; PRT; 1950 AA.
 AC Q26080;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Myosin heavy chain.
 OS Placopecten magellanicus (Sea scallop).
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Pectinoidea;
 OC Pectinoidea; Pectinidae; Placopecten.
 OX NCBI_TaxID=6577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Adductor muscle;
 RA Perreault-Micale C., Kalabokis V., Nyitray L., Szent-Gyorgyi A.G.;
 RT "Sequence variations in the surface loop near the nucleotide binding
 RT site modulate the ATP turnover rates of molluscan myosins.";
 RL J. Muscle Res. Cell Motil. 0:0-0(1996).
 DR EMBL; U59295; AAB03661.1; -.
 DR HSSP; P24733; IWDC.
 DR InterPro; IPR000048; IQ region.
 DR InterPro; IPR001609; myosin_head.
 DR InterPro; IPR004009; Myosin_N.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 13:18:03 ; Search time 11 Seconds
(without alignments)
166.731 Million cell updates/sec

Title: US-10-025-598-2

Perfect score: 190

Sequence: 1 MDWLKARVEQELQALEARGTDSNAELRAMEAKLKAEIQK 39

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60.5	31.8	794	1 HMMR_MOUSE	Q00547 mus musculus
2	60.5	31.8	1549	1 SMC4_CAEEL	Q20060 caenorhabdi
3	59	31.1	316	1 APE_BOVIN	Q03247 bos taurus
4	58.5	30.8	224	1 CTH8_HUMAN	Q9h444 homo sapien
5	58.5	30.8	224	1 CTH8_MOUSE	Q9d8b3 mus musculus
6	58.5	30.8	624	1 PBPD_BACSU	P40750 bacillus su
7	58	30.5	177	1 IF3_ANASP	Q8yne3 anabaena sp
8	57	30.0	1243	1 SMC4_MICAR	Q9era5 microtus ar
9	56	29.5	107	1 APE_SAISC	Q28995 saimiri sci
10	56	29.5	574	1 LAMC_MOUSE	P11516 mus musculus
11	56	29.5	665	1 LAMA_MOUSE	P48678 mus musculus
12	56	29.5	665	1 LAMA_RAT	P48679 rattus norv
13	55.5	29.2	186	1 TATB_HAEIN	P57047 haemophilus
14	55.5	29.2	577	1 CST2_HUMAN	P33240 homo sapien
15	55.5	29.2	595	1 MERL_HUMAN	P35240 homo sapien
16	55.5	29.2	596	1 MERL_MOUSE	P46662 mus musculus
17	55	28.9	298	1 APE_CAVPO	P23529 cavia porce
18	55	28.9	664	1 LAMA_HUMAN	P02545 homo sapien
19	55	28.9	864	1 STA2_PIG	O02799 sus scrofa
20	54.5	28.7	222	1 C134_HUMAN	Q9by43 homo sapien
21	54.5	28.7	854	1 DIS1_HUMAN	Q9nri5 homo sapien
22	54	28.4	171	1 IF3_THETH	Q9acj8 thermus the
23	54	28.4	182	1 IF3_SYNEL	Q8di98 synechococc
24	54	28.4	454	1 AK3_BACSU	P94417 bacillus su
25	54	28.4	970	1 TRAI_COMTE	Q04222 comamonas t
26	54	28.4	2442	1 CEP2_HUMAN	Q9bv73 homo sapien
27	53.5	28.2	1976	1 MYHA_BOVIN	Q27991 bos taurus
28	53.5	28.2	1976	1 MYHA_HUMAN	P35580 homo sapien
29	53	27.9	151	1 WTAP_HUMAN	Q15007 homo sapien
30	53	27.9	1008	1 MOG4_CAEEL	O45244 caenorhabdi
31	53	27.9	1690	1 C190_DROME	Q9vie5 drosophila
32	53	27.9	2349	1 TPR_HUMAN	P12270 homo sapien
33	52.5	27.6	169	1 MEMG_METCA	P11987 methylococc

RESULT 1				
HMMR_MOUSE				
ID	HMMR_MOUSE	STANDARD;	PRT;	794 AA.
AC	Q00547;			
DT	01-OCT-1996	(Rel. 34, Created)		
DT	30-MAY-2000	(Rel. 39, Last sequence update)		
DT	28-FEB-2003	(Rel. 41, Last annotation update)		
DE	Hyaluronan mediated motility receptor (Intracellular hyaluronic acid binding protein) (Receptor for hyaluronan-mediated motility).			
GN	HMMR OR IHABP OR RHAMM.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.			
RC	TISSUE=Lung;			
RX	MEDLINE=98264863; PubMed=9601097;			
RA	Hofmann M., Fieber C., Assmann V., Goettlicher M., Sleeman J.,			
RA	Plug R., Howells N., von Stein O., Ponta H., Herrlich P.;			
RT	"Identification of IHABP, a 95 kDa intracellular hyaluronate binding protein."			
RL	J. Cell Sci. 111:1673-1684(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Zhao Y., Zhang S., Turley E.;			
RL	Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE OF 1-183 FROM N.A., AND TISSUE SPECIFICITY.			
RC	STRAIN=129/Sv;			
RX	MEDLINE=99107769; PubMed=9889313;			
RA	Fieber C., Plug R., Sleeman J., Dall P., Ponta H., Hofmann M.;			
RT	"Characterization of the murine gene encoding the intracellular hyaluronan receptor IHABP."			
RL	Gene 226:41-50(1999).			
RN	[4]			
RP	SEQUENCE OF 164-794 FROM N.A., AND ALTERNATIVE SPLICING.			
RC	STRAIN=BALB/c; TISSUE=Fibroblast;			
RX	MEDLINE=96011639; PubMed=7590272;			
RA	Entwistle J., Zhang S., Yang B., Wong C., Li Q., Hall C.L., A J.,			
RA	Mowat M., Greenberg A.H., Turley E.A.;			
RT	"Characterization of the murine gene encoding the hyaluronan receptor RHAMM."			
RL	Gene 163:233-238(1995).			
RN	[5]			
RP	SEQUENCE OF 318-794 FROM N.A., AND CHARACTERIZATION.			
RC	STRAIN=BALB/c;			
RX	MEDLINE=92299690; PubMed=1376732;			
RA	Hardwick C., Hoare K., Owens R., Hohn H.P., Hook M., Moore D.,			
RA	Cripps V., Austen L., Nance D.M., Turley E.A.;			
RT	"Molecular cloning of a novel hyaluronan receptor that mediates tumor cell motility."			
RL	J. Cell Biol. 117:1343-1350(1992).			
RN	[6]			
RP	CHARACTERIZATION.			
RX	MEDLINE=94308286; PubMed=7518470;			

P25386 saccharomyc
Q28502 macaca mula
Q9kpw1 vibrio chol
P72874 synechocyst
P10517 macaca fasc
P05770 papio anubi
P13925 streptococc
Q15025 homo sapien
P13648 gallus gall
Q09863 schizosacch
P24733 aequipecten
P35749 homo sapien

ALIGNMENTS

10/025,598
Search Notes
2/21/04


```
CC -!- SIMILARITY: Belongs to the IF-3 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP003597; BAB76322.1; ALT_INIT.
CC HAMAP; MF 00080; -; 1.
CC InterPro; IPR001288; IF3.
CC Pfam; PF00707; IF3_C; 1.
CC Pfam; PF05198; IF3_N; 1.
CC ProDom; PD002880; IF3; 1.
CC TIGRFAMs; TIGR00168; infC; 1.
CC PROSITE; PS00938; IF3; 1.
KW Initiation factor; Protein biosynthesis; Complete proteome.
SQ SEQUENCE 177 AA; 20750 MW; 9C1585FF62590EFF CRC64;

Query Match 30.5%; Score 58; DB 1; Length 177;
Best Local Similarity 36.1%; Pred.No. 7.9;
Matches 13; Conservative 9; Mismatches 14; Indels 0; Gaps 0;

Qy 1 MDWLKARVEQLQALEARGTDSNAELRAEAKLKAE 36
Db 68 MDYGYKFEQEKAREARKKQHTADVKEVMRYKIE 103

RESULT 8
SMC4_MICAR STANDARD; PRT; 1243 AA.
ID SMC4_MICAR Q9ERA5; Q9ERA1;
AC Q9ERA5; Q9ERA1;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Structural maintenance of chromosomes 4-like 1 protein (Chromosome-
DE associated polypeptide C) (XCAP-C homolog) (Fragment).
GN SMC4L1 OR SMC4 OR CAPC.
OS Microtus arvalis (Common vole) (Field vole).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Arvicolinae;
OC Microtus.
OC NCBI_TaxID=47230;
RN [1]
RP SEQUENCE FROM N.A.
RA Pavlova S.V., Nesterova T.B., Zakian S.M.;
RT "SMC genes from common vole Microtus arvalis.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Central component of the condensin complex, a complex
CC required for conversion of interphase chromatin into mitotic-like
CC condense chromosomes. The condensin complex probably introduces
CC positive supercoils into relaxed DNA in the presence of type I
CC topoisomerases and converts nicked DNA into positive knotted forms
CC in the presence of type II topoisomerases (By similarity).
CC -!- SUBUNIT: Forms an heterodimer with SMC2L1. Component of the
CC condensin complex, which contains the SMC2L1 and SMC4L1
CC heterodimer, and three non SMC subunits that probably regulate the
CC complex: BRRN1/CAPH, CNAPl/CAPD2 and CAPG (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic. In interphase
CC cells, the majority of the condensin complex is found in the
CC cytoplasm, while a minority of the complex is associated with
CC chromatin. A subpopulation of the complex however remains
CC associated with chromosome foci in interphase cells. During
CC mitosis, most of the condensin complex is associated with the
CC chromatin. At the onset of prophase, the regulatory subunits of
CC the complex are phosphorylated by CDC2, leading to condensin's
CC association with chromosome arms and to chromosome condensation.
CC Dissociation from chromosomes is observed in late telophase (By
CC similarity).
CC -!- DOMAIN: The hinge domain, which separates the large intramolecular
CC coiled coil regions, allows the heterodimerization with SMC2L1,
```

```
CC forming a V-shaped heterodimer (By similarity).
CC -!- SIMILARITY: Belongs to the SMC family. SMC4 subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ299713; CAC09583.1; -.
CC EMBL; AJ299717; CAC09587.1; -.
CC InterPro; IPR003439; ABC transporter.
CC InterPro; IPR003405; SMC_C.
CC InterPro; IPR003395; SMC_N.
CC Pfam; PF02483; SMC_C; 1.
CC Pfam; PF02463; SMC_N; 1.
KW DNA condensation; Mitosis; Cell cycle; ATP-binding; Coiled coil;
KW Nuclear protein.
FT NON_TER 1
FT NP_BIND 68 75 ATP (POTENTIAL).
FT DOMAIN 225 546 COILED COIL (POTENTIAL).
FT DOMAIN 547 720 FLEXIBLE HINGE.
FT DOMAIN 721 975 COILED COIL (POTENTIAL).
FT DOMAIN 1062 1090 COILED COIL (POTENTIAL).
FT DOMAIN 1146 1181 ALA/ASP-RICH (DA-BOX).
FT DOMAIN 21 26 POLY-PRO.
FT CONFLICT 241 241 N -> K (IN REF. 1; CAC09587).
FT CONFLICT 757 757 E -> R (IN REF. 1; CAC09587).
SQ SEQUENCE 1243 AA; 142341 MW; B621BD6D954DA7F9 CRC64;

Query Match 30.0%; Score 57; DB 1; Length 1243;
Best Local Similarity 39.4%; Pred.No. 69;
Matches 13; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

Qy 7 RVEQLQALEARGTDSNAELRAEAKLKAEIQK 39
Db 898 RTEKIKDTEKETNDLKAEKAEVIRK 930

RESULT 9
APE_SAISC STANDARD; PRT; 107 AA.
ID APE_SAISC Q28995;
AC Q28995;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Apolipoprotein E (Apo-E) (Fragment).
GN APOE.
OS Saimiri sciureus (Common squirrel monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Saimiri.
OC NCBI_TaxID=9521;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96225955; PubMed=8635577;
RA Moralli L., Wei L., Amorim A., McDermid J., Abee C.R., Frangione B.,
RA Walker L.C., Levy E.;
RT "Cerebrovascular amyloidosis in squirrel monkeys and rhesus monkeys:
RT apolipoprotein E genotype.";
RL FEBS Lett. 379:132-134(1996).
CC -!- FUNCTION: APO-E MEDIATES BINDING, INTERNALIZATION, AND CATABOLISM
CC OF LIPOPROTEIN PARTICLES. IT CAN SERVE AS A LIGAND FOR THE LDL(APO
CC B/E) RECEPTOR AND FOR THE SPECIFIC APO-E RECEPTOR (CHYLOMICRON
CC REMNANT) OF HEPATIC TISSUES.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Secreted in plasma.
CC -!- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```



```
FT DOMAIN 81 218 COIL 1B.
FT DOMAIN 219 242 LINKER 2.
FT DOMAIN 243 383 COIL 2.
FT SITE 325 325 STUTTER (BY SIMILARITY).
FT SITE 266 266 HEPTAD CHANGE OF PHASE.
FT SITE 330 330 HEPTAD CHANGE OF PHASE.
FT DOMAIN 417 422 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD_RES 392 392 PHOSPHORYLATION (BY CDC2) (PARTIAL).
FT MOD_RES 407 407 PHOSPHORYLATION (PARTIAL).
FT MOD_RES 409 409 PHOSPHORYLATION (PARTIAL).
FT VARSPLIC 1 112 Missing (in isoform C).
FT VARSPLIC 113 118 /FTID=VSP 002471.
FT VARSPLIC 118 118 FKELKA -> MGNAEG (in isoform C).
FT CONFLICT 118 119 /FTID=VSP 002472.
FT CONFLICT 401 401 AR -> VC (IN REF. 1).
FT CONFLICT 439 440 R -> P (IN REF. 1).
FT CONFLICT 439 440 RV -> WL (IN REF. 1).
SQ SEQUENCE 574 AA; 65446 MW; A736DFICEDB65BE CRC64;

Query Match 29.5%; Score 56; DB 1; Length 574;
Best Local Similarity 37.9%; Pred. No. 42;
Matches 11; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

QY 6 ARVEQELQALEARGTDSNAELRAMEAKLK 34
Db 107 SKVREEFKELKARNTKKEGDLAAQARLK 135

RESULT 11
LAMA_MOUSE
ID LAMA_MOUSE STANDARD; PRT; 665 AA.
AC P48678; P97859;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Lamin A.
GN LMNA OR LMN1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95300954; PubMed=7781761;
RA Nakajima N., Abe K.;
RT "Genomic structure of the mouse A-type lamin gene locus encoding
RT somatic and germ cell-specific lamins.";
RL FEBS Lett. 365:108-114(1995).
RN [2]
RP SEQUENCE OF 1-568 FROM N.A.
RX MEDLINE=89247440; PubMed=2719959;
RA Riedel W., Werner D.;
RT "Nucleotide sequence of the full-length mouse lamin C cDNA and its
RT deduced amino-acid sequence.";
RL Biochim. Biophys. Acta 1008:119-122(1989).
RN [3]
RP SEQUENCE OF 235-665 FROM N.A.
RX MEDLINE=93144345; PubMed=7916626;
RA Nakajima N., Sado T.;
RT "Nucleotide sequence of a mouse lamin A cDNA and its deduced amino
RT acid sequence.";
RL Biochim. Biophys. Acta 1171:311-314(1993).
CC -!- FUNCTION: LAMINS ARE COMPONENTS OF THE NUCLEAR LAMINA, A FIBROUS
CC LAYER ON THE NUCLEOPLASMIC SIDE OF THE INNER NUCLEAR MEMBRANE,
CC WHICH IS THOUGHT TO PROVIDE A FRAMEWORK FOR THE NUCLEAR ENVELOPE
CC AND MAY ALSO INTERACT WITH CHROMATIN.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=A;
CC IsoId=P48678-1; Sequence=Displayed;
CC Note=Isoforms A and C are present in equal amounts in the
CC lamina of mammals;
```

```
CC Name=B;
CC IsoId=P11516-1; Sequence=External;
CC Name=C;
CC IsoId=P11516-2; Sequence=External;
CC Note=Isoforms A and C are present in equal amounts in the
CC lamina of mammals;
CC -!- PTM: INCREASED PHOSPHORYLATION OF THE LAMINS OCCURS BEFORE
CC ENVELOPE DISINTEGRATION AND PROBABLY PLAYS A ROLE IN REGULATING
CC LAMIN ASSOCIATIONS.
CC -!- MISCELLANEOUS: THE STRUCTURAL INTEGRITY OF THE LAMINA IS STRICTLY
CC CONTROLLED BY THE CELL CYCLE, AS SEEN BY THE DISINTEGRATION AND
CC FORMATION OF THE NUCLEAR ENVELOPE IN PROPHASE AND TELOPHASE,
CC RESPECTIVELY.
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 582
CC ONWARD AND IS LONGER (702 AA) DUE TO A FRAMESHIFT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D49733; BAA08569.1; -.
DR EMBL; D49732; BAA08569.1; JOINED.
DR EMBL; D49734; BAA08569.1; JOINED.
DR EMBL; D49735; BAA08569.1; JOINED.
DR EMBL; D13181; BAA02476.1; -.
DR MGD; MGI:96794; Lmna.
DR GO; GO:0005638; C:lamin filament; IDA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR001322; IF_tail.
DR Pfam; PF00038; filament; 1.
DR Pfam; PF00932; IF_tail; 1.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Nuclear protein; Phosphorylation;
KW Alternative splicing.
FT DOMAIN 1 33 HEAD.
FT DOMAIN 34 383 ROD.
FT DOMAIN 384 665 TAIL.
FT DOMAIN 34 70 COIL 1A.
FT DOMAIN 71 80 LINKER 1.
FT DOMAIN 81 218 COIL 1B.
FT DOMAIN 219 242 LINKER 2.
FT DOMAIN 243 383 COIL 2.
FT SITE 325 325 STUTTER (BY SIMILARITY).
FT DOMAIN 417 422 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT LIPID 662 662 FARNESYL (BY SIMILARITY).
FT CONFLICT 118 119 AR -> VC (IN REF. 2).
FT CONFLICT 401 401 R -> P (IN REF. 2).
FT CONFLICT 439 440 RV -> WL (IN REF. 2).
FT CONFLICT 623 623 A -> V (IN REF. 1).
SQ SEQUENCE 665 AA; 74209 MW; 5434E6D60A04CB15 CRC64;

Query Match 29.5%; Score 56; DB 1; Length 665;
Best Local Similarity 37.9%; Pred. No. 48;
Matches 11; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

QY 6 ARVEQELQALEARGTDSNAELRAMEAKLK 34
Db 107 SKVREEFKELKARNTKKEGDLAAQARLK 135

RESULT 12
LAMA_RAT
ID LAMA_RAT STANDARD; PRT; 665 AA.
AC P48679;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Lamin A.
```

GN LMNA OR LMN1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93050186; PubMed=1426247;
RA Ozaki T., Sakiyama S.;
RT "Lamin A gene expression is specifically suppressed in
RT v-src-transformed cells."
RL FEBS Lett. 312:165-168(1992).
RN [2]
RP SEQUENCE OF 26-663 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RA Jonnalagadda V.S., Parnaik V.K.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: LAMINS ARE COMPONENTS OF THE NUCLEAR LAMINA, A FIBROUS
CC LAYER ON THE NUCLEOPLASMIC SIDE OF THE INNER NUCLEAR MEMBRANE,
CC WHICH IS THOUGHT TO PROVIDE A FRAMEWORK FOR THE NUCLEAR ENVELOPE
CC AND MAY ALSO INTERACT WITH CHROMATIN.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Lamin A;
CC IsoId=P48679-1; Sequence=Displayed;
CC Note=Isoforms A and C are present in equal amounts in the lamina
CC of mammals;
CC Name=Lamin C;
CC IsoId=P48679-2; Sequence=Not described;
CC Note=Isoforms A and C are present in equal amounts in the lamina
CC of mammals;
CC -!- PTM: INCREASED PHOSPHORYLATION OF THE LAMINS OCCURS BEFORE
CC ENVELOPE DISINTEGRATION AND PROBABLY PLAYS A ROLE IN REGULATING
CC LAMIN ASSOCIATIONS.
CC -!- MISCELLANEOUS: THE STRUCTURAL INTEGRITY OF THE LAMINA IS STRICTLY
CC CONTROLLED BY THE CELL CYCLE, AS SEEN BY THE DISINTEGRATION AND
CC FORMATION OF THE NUCLEAR ENVELOPE IN PROPHASE AND TELOPHASE,
CC RESPECTIVELY.
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN IN THE N- AND C-
CC TERMINAL DUE TO FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X66870; CAA47342.1; ALT_FRAME.
DR EMBL; X76297; CAA53945.1; -.
DR InterPro; IPR001664; IF.
DR InterPro; IPR001322; IF_tail.
DR Pfam; PF00038; filament; 1.
DR Pfam; PF00932; IF_tail; 1.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Nuclear protein; Lipoprotein;
KW Prenylation; Phosphorylation; Alternative splicing.
FT DOMAIN 1 33 HEAD.
FT DOMAIN 34 383 ROD.
FT DOMAIN 384 665 TAIL.
FT DOMAIN 34 70 COIL 1A.
FT DOMAIN 71 80 LINKER 1.
FT DOMAIN 81 218 COIL 1B.
FT DOMAIN 219 242 LINKER 2.
FT DOMAIN 243 383 COIL 2.
FT SITE 325 325 STUTTER (BY SIMILARITY).
FT DOMAIN 417 422 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT LIPID 662 662 FARNESYL (BY SIMILARITY).
FT CONFLICT 470 470 K -> R (IN REF. 1).
FT CONFLICT 524 524 T -> S (IN REF. 1).

FT CONFLICT 584 584 R -> P (IN REF. 1).
FT CONFLICT 606 606 A -> P (IN REF. 1).
SQ SEQUENCE 665 AA; 74323 MW; 9CC553005C8534E4 CRC64;
Query Match 29.5%; Score 56; DB 1; Length 665;
Best Local Similarity 37.9%; Pred. No. 48;
Matches 11; Conservative 8; Mismatches 10; Indels 0; Gaps 0;
QY 6 ARVEQELQALEARGTDSNAELRAMEAKLK 34
Db 107 SKVREFKELKARNTKKEGDLAAQARLK 135
RESULT 13
TATB_HAEIN STANDARD; PRT; 186 AA.
AC P57047; P44559;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sec-independent protein translocase protein tatB homolog.
GN TATB OR HI0187.1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Keiley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -!- FUNCTION: Required for correct localization of precursor proteins
CC bearing signal peptides with the twin arginine conserved motif
CC S/T-R-R-X-F-L-K. This sec-independent pathway is termed TAT for
CC twin-arginine translocation system. This system mainly transports
CC proteins with bound cofactors that require folding prior to export
CC (By similarity).
CC -!- SUBCELLULAR LOCATION: Inner-membrane bound (Probable).
CC -!- SIMILARITY: Belongs to the tatB family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32704; -. NOT_ANNOTATED_CDS.
DR TIGR; HI0187.1; -.
DR HAMAP; MF_00237; -. 1.
DR InterPro; IPR003998; TatB.
DR PRINTS; PR01506; TATBPROTEIN.
DR TIGRFAMs; TIGR01410; tatB; 1.
KW Transport; Protein transport; Translocation; Transmembrane;
KW Inner membrane; Complete proteome.
FT TRANSMEM 1 21 POTENTIAL.
SQ SEQUENCE 186 AA; 20430 MW; 7FD2166488CA057F CRC64;
Query Match 29.2%; Score 55.5; DB 1; Length 186;
Best Local Similarity 27.7%; Pred. No. 16;
Matches 18; Conservative 8; Mismatches 12; Indels 27; Gaps 2;

RL Cell 72:791-800(1993).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=93281181; PubMed=8379998;
RA Rouleau G.A., Merel P., Lutchman M., Sanson M., Zucman J.,
RA Marineau C., Hoang-Xuan K., Demczuk S., Desmaze C., Plougastel B.,
RA Pulst S., Lenoir G., Bijlsma E., Fashold R., Dumanski J., de Jong P.,
RA Parry D., Eldridge R., Aurias A., Delattre O., Thomas G.;
RT "Alteration in a new gene encoding a putative membrane-organizing
RL protein causes neuro-fibromatosis type 2.";
RL Nature 363:515-521(1993).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX MEDLINE=99036680; PubMed=9817927;
RA Zucman-Rossi J., Legoux P., Der Sarkissian H., Cheret G., Sor F.,
RA Bernardi A., Cazes L., Giraud S., Lenoir G., Thomas G.;
RT "NF2 gene in neurofibromatosis type 2 patients.";
RL Hum. Mol. Genet. 7:2095-2101(1998).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORMS 7; 9 AND 10), AND SUBCELLULAR LOCATION.
RX MEDLINE=99330566; PubMed=10401006;
RA Schmucker B., Tang Y., Kressel M.;
RT "Novel alternatively spliced isoforms of the neurofibromatosis type 2
RL tumor suppressor are targeted to the nucleus and cytoplasmic
RT granules.";
RL Hum. Mol. Genet. 8:1561-1570(1999).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 3; 4; 5; 6 AND 8).
RX MEDLINE=21686156; PubMed=11827459;
RA Chang L.-S., Akhmet'yeva E.M., Wu Y., Zhu L., Welling D.B.;
RA "Multiple transcription initiation sites, alternative splicing, and
RT differential polyadenylation contribute to the complexity of human
RT neurofibromatosis 2 transcripts.";
RL Genomics 79:63-76(2002).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 4).
RC TISSUE=Lung, and Skin;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zebberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RL human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [7]
RP REVIEW.
RA Marineau C., Merel P., Rouleau G.A., Thomas G.;
RT "The gene of neurofibromatosis type 2.";
RL Medecine/Sciences 11:35-42(1995).
RN [8]
RP VARIANT NF2 TYR-220.
RX MEDLINE=94047514; PubMed=8230593;
RA Maccollin M.M., Mohny T., Trofatter J.A., Wertelecki W., Ramesh V.,
RA Gusella J.F.;
RT "DNA diagnosis of neurofibromatosis 2. Altered coding sequence of the
RT merlin tumor suppressor in an extended pedigree.";
RL JAMA 270:2316-2320(1993).
RN [9]
RP VARIANT NF2 PHE-96 DEL.

RX MEDLINE=94311312; PubMed=7913580;
RA Maccollin M.M., Ramesh V., Jacoby L.B., Louis D.N., Rubio M.-P.,
RA Pulaski K., Trofatter J.A., Short M.P., Bove C., Eldridge R.,
RA Parry D.M., Gusella J.F.;
RT "Mutational analysis of patients with neurofibromatosis 2.";
RL Am. J. Hum. Genet. 55:314-320(1994).
RN [10]
RP VARIANT ARG-46.
RX MEDLINE=94272478; PubMed=8004107;
RA Irving R.M., Moffat D.A., Hardy D.G., Barton D.E., Xuereb J.H.,
RA Maher E.R.;
RT "Somatic NF2 gene mutations in familial and non-familial vestibular
RL schwannoma.";
RL Hum. Mol. Genet. 3:347-350(1994).
RN [11]
RP VARIANTS MET-219 AND CYS-418.
RX MEDLINE=94282032; PubMed=8012353;
RA Jacoby L.B., Maccollin M.M., Louis D.N., Mohny T., Rubio M.-P.,
RA Pulaski K., Trofatter J.A., Kley N., Seizinger B.R., Ramesh V.,
RA Gusella J.F.;
RT "Exon scanning for mutation of the NF2 gene in schwannomas.";
RL Hum. Mol. Genet. 3:413-419(1994).
RN [12]
RP VARIANTS NF2 SER-62; GLY-106 AND MET-352.
RX MEDLINE=94362682; PubMed=8081368;
RA Bourn D., Carter S.A., Mason S., Gareth D., Evans R., Strachan T.;
RT "Germline mutations in the neurofibromatosis type 2 tumour suppressor
RT gene.";
RL Hum. Mol. Genet. 3:813-816(1994).
RN [13]
RP VARIANT GLY-79.
RX MEDLINE=95038740; PubMed=7951231;
RA Sainz J., Huynh D.P., Figueroa K., Ragge N.K., Baser M.E., Pulst S.M.;
RT "Mutations of the neurofibromatosis type 2 gene and lack of the gene
RT product in vestibular schwannomas.";
RL Hum. Mol. Genet. 3:885-891(1994).
RN [14]
RP VARIANTS PHE-273 AND ILE-364.
RX MEDLINE=94214500; PubMed=8162073;
RA Bianchi A.B., Hara T., Ramesh V., Gao J., Klein Szanto A.J., Morin F.,
RA Menon A.G., Trofatter J.A., Gusella J.F., Seizinger B.R., Kley N.;
RT "Mutations in transcript isoforms of the neurofibromatosis 2 gene in
RT multiple human tumour types.";
RL Nat. Genet. 6:185-192(1994).
RN [15]
RP VARIANT NF2 PRO-535.
RX MEDLINE=95395825; PubMed=7666400;
RA Evans D.G.R., Bourn D., Wallace A., Ramsden R.T., Mitchell J.D.,
RA Strachan T.;
RT "Diagnostic issues in a family with late onset type 2
RT neurofibromatosis.";
RL J. Med. Genet. 32:470-474(1995).
RN [16]
RP VARIANT NF2 PRO-538.
RX MEDLINE=96163499; PubMed=8566958;
RA Kluwe L., Mautner V.-F.;
RT "A missense mutation in the NF2 gene results in moderate and mild
RT clinical phenotypes of neurofibromatosis type 2.";
RL Hum. Genet. 97:224-227(1996).
RN [17]
RP VARIANTS PHE-96 DEL; ILE-117; PHE-119 DEL; 122-VAL--GLU-129 DEL AND
RP PHE-339.
RX MEDLINE=96209910; PubMed=8655144;
RA de Vitis L.R., Tedde A., Vitelli F., Ammannati F., Mennonna P.,
RA Bigozzi U., Montali E., Papi L.;
RT "Screening for mutations in the neurofibromatosis type 2 (NF2) gene
RT in sporadic meningiomas.";
RL Hum. Genet. 97:632-637(1996).
RN [18]
RP VARIANT NF2 ARG-234.
RX MEDLINE=99192373; PubMed=10090912;
RA Baser M.E., Kluwe L., Mautner V.-F.;
RT "Germ-line NF2 mutations and disease severity in neurofibromatosis

RT type 2 patients with retinal abnormalities.";
RL Am. J. Hum. Genet. 64:1230-1233(1999).
CC -!- FUNCTION: Probably acts as a membrane stabilizing protein.
CC -!- SUBCELLULAR LOCATION: In a fibroblastic cell line, isoforms 1 and
CC 10 are found homogeneously distributed over the entire cell, with a
CC particularly strong staining in ruffling membranes and filopodia.
CC Isoform 10 is also found in the nucleus. Isoforms 7, 9 and 10 are
CC observed in cytoplasmic granules concentrated in a perinuclear
CC location. Isoforms 7 and 9 are absent from ruffling membranes and
CC filopodia.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=10;
CC Name=1; Synonyms=1;
CC IsoId=P35240-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P35240-2; Sequence=VSP_000492;
CC Name=3; Synonyms=11;
CC IsoId=P35240-3; Sequence=VSP_007050, VSP_007051;
CC Name=4; Synonyms=delete2/3;
CC IsoId=P35240-4; Sequence=VSP_007041, VSP_007050, VSP_007051;
CC Name=5; Synonyms=delete3;
CC IsoId=P35240-5; Sequence=VSP_007042, VSP_007050, VSP_007051;
CC Name=6; Synonyms=delete2;
CC IsoId=P35240-6; Sequence=VSP_007040, VSP_007050, VSP_007051;
CC Name=7; Synonyms=MER150;
CC IsoId=P35240-7; Sequence=VSP_007045, VSP_007046;
CC Name=8;
CC IsoId=P35240-8; Sequence=VSP_007048, VSP_007050, VSP_007051;
CC Name=9; Synonyms=MER162;
CC IsoId=P35240-9; Sequence=VSP_007044;
CC Name=10; Synonyms=MER151;

Query Match 29.2%; Score 55.5; DB 1; Length 595;
Best Local Similarity 42.4%; Pred. No. 49;
Matches 14; Conservative 6; Mismatches 12; Indels 1; Gaps 1;

Qy 6 ARVEQLQALEARGTDSNAELRAMEAK-LKAEI 37
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 403 AAEQEMQRIKATAIRTEEEKRLMEQKVLAEAV 435

Search completed: January 30, 2004, 13:21:24
Job time : 12 secs

hypothetical protein T20K18.120 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 22-Oct-1999
C;Accession: T06635
R;Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; Mew
submitted to the Protein Sequence Database, April 1999
A;Reference number: Z15790
A;Accession: T06635
A;Molecule type: DNA
A;Residues: 1-909 <BEV>
A;Cross-references: EMBL:AL049640; GSPDB:GN00062; ATSP:T20K18.120
A;Experimental source: cultivar Columbia; BAC clone T20K18
C;Genetics:
A;Gene: ATSP:T20K18.120
A;Map position: 4
A;Introns: 205/2; 686/3; 740/3; 772/2; 808/3; 838/3; 897/3

Query Match 30.8%; Score 58.5; DB 2; Length 909;
Best Local Similarity 51.6%; Pred. No. 52;
Matches 16; Conservative 5; Mismatches 9; Indels 1; Gaps 1;

Qy 6 ARVEQELQALEARGTDSNAELRAMEAKLKAE 36
Db 605 AREKAEKAAAEAR-ERANAEREKEAKVRAE 634

RESULT 8
AG2383
translation initiation factor IF-3 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AG2383
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi
Nakazaki, N.; Shimpko, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AG2383
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-204 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB76322.1; PID:g17133760; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: infC
C;Superfamily: translation initiation factor IF-3

Query Match 30.5%; Score 58; DB 2; Length 204;
Best Local Similarity 36.1%; Pred. No. 13;
Matches 13; Conservative 9; Mismatches 14; Indels 0; Gaps 0;

Qy 1 MDWLKARVEQELQALEARGTDSNAELRAMEAKLKAE 36
Db 95 MDYGYKFEQEKAREARKKQHTADVKEVKMYKIE 130

RESULT 9
D86303
F17F16.1 protein - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 27-Nov-2001
C;Accession: D86303
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: D86303
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2142 <STO>
A;Cross-references: GB:AE005172; NID:g9954728; PIDN:AAG09081.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1

Query Match 30.5%; Score 58; DB 2; Length 2142;
Best Local Similarity 31.4%; Pred. No. 1.4e+02;
Matches 11; Conservative 12; Mismatches 12; Indels 0; Gaps 0;

Qy 5 KARVEQELQALEARGTDSNAELRAMEAKLKAEIQK 39
Db 1451 KRKIYKDL SAVQAQERKANYEMRTLKQKLRKSILK 1485

RESULT 10
E90243
hypothetical protein SSO0922 [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C;Accession: E90243
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-y
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: E90243
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-122 <KUR>
A;Cross-references: GB:AE006641; NID:g13814104; PIDN:AAK41204.1; GSPDB:GN00155
C;Genetics:
A;Gene: SSO0922

Query Match 30.3%; Score 57.5; DB 2; Length 122;
Best Local Similarity 31.6%; Pred. No. 8.6;
Matches 12; Conservative 11; Mismatches 8; Indels 7; Gaps 1;

Qy 2 DWLKARVEQELQALEARGTDSNAELRAMEAKLKAEIQK 39
Db 9 DMVKSMEKEINRME-----KEFRRIESEIKEIEK 39

RESULT 11
T45983
hypothetical protein F9D24.140 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T45983
R;D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Lemcke, K.;
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z23011
A;Accession: T45983
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-153 <DAN>
A;Cross-references: EMBL:AL137081
A;Experimental source: cultivar Columbia; BAC clone F9D24
C;Genetics:
A;Map position: 3
A;Introns: 19/3
A;Note: F9D24.140

Query Match 30.3%; Score 57.5; DB 2; Length 153;
Best Local Similarity 32.7%; Pred. No. 11;
Matches 16; Conservative 9; Mismatches 11; Indels 13; Gaps 2;

Qy 1 MDWLKARVEQ---ELQALEARG-----TDSNAELRAMEAKLKAE 36
:||||| |:| : : :|:|

Db 88 LDWLKRLBELCVKKKMEARGARMRELDGMIVEQRRVLWALETELKNE 136

RESULT 12

C75610

hypothetical protein - Deinococcus radiodurans (strain R1)

C;Species: Deinococcus radiodurans

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000

C;Accession: C75610

R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A;Reference number: A75250; MUID:20036896; PMID:10567266

A;Accession: C75610

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-259 <WHI>

A;Cross-references: GB:AE001862; GB:AE001825; NID:G6460468; PIDN:AAF12313.1; PID:G646060

A;Experimental source: strain R1

C;Genetics:

A;Gene: DRA0144

A;Map position: 2

C;Superfamily: Deinococcus radiodurans hypothetical protein DRA0144

Query Match 30.3%; Score 57.5; DB 2; Length 259;

Best Local Similarity 40.9%; Pred. No. 19;

Matches 18; Conservative 6; Mismatches 9; Indels 11; Gaps 2;

Qy 7 RVEQELQALE-----ARGTD--SNAELRAMEAKLKAEIQK 39

|||: ||||| | : | ||||| : ||| : |||

Db 207 RVQENQALETFMDKEYSKTAKPVDMSAEELRASQKQLQEFLOK 250

RESULT 13

S49245

19K protein - soil-borne wheat mosaic virus (isolate OK1-1)

C;Species: soil-borne wheat mosaic virus, SBWMV

A;Variety: isolate OK1-1

C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 05-Nov-1999

C;Accession: S49245

R;Chen, J.; MacFarlane, S.; Michael, T.; Wilson, A.

submitted to the EMBL Data Library, September 1994

A;Description: Deletion formation in the readthrough-protein gene from SBWMV RNA2.

A;Reference number: S49244

A;Accession: S49245

A;Molecule type: genomic RNA

A;Residues: 1-174 <CHE>

A;Cross-references: EMBL:X81639; NID:G547536; PIDN:CAA57295.1; PID:G547538

A;Experimental source: isolate OK1-1; strain Oklahoma

C;Superfamily: soil-borne wheat mosaic virus hypothetical 19K protein

Query Match 29.7%; Score 56.5; DB 2; Length 174;

Best Local Similarity 31.5%; Pred. No. 16;

Matches 17; Conservative 7; Mismatches 11; Indels 19; Gaps 2;

Qy 3 WLKARVEQ---ELQALEARGTDSNAELRAMEA-----KLKAEI 37

| : : | ||||| | : : | |

Db 84 WRHAQLRSLNAELDALEAREESLRQIKALSAGDHCPAVLAYVPKLTCLKAEV 137

RESULT 14

E87478

GGDEF family protein [imported] - Caulobacter crescentus

C;Species: Caulobacter crescentus

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C;Accession: E87478

R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J. B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A;Title: Complete Genome Sequence of Caulobacter crescentus.

A;Reference number: A87249; MUID:21173698; PMID:11259647

A;Accession: E87478

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-353 <STO>

A;Cross-references: GB:AE005673; NID:gl3423291; PIDN:AAK23825.1; GSPDB:GN00148

C;Genetics:

A;Gene: CCL850

Query Match 29.5%; Score 56; DB 2; Length 353;

Best Local Similarity 46.4%; Pred. No. 38;

Matches 13; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

Qy 7 RVEQELQALEARGTDSNAELRAMEAKLK 34

|||: ||||| | : | ||||| : ||| : |||

Db 147 RVHKENQSLARLADSTAEVERLFEHLE 174

RESULT 15

G71490

hypothetical protein CT622 - Chlamydia trachomatis (serotype D, strain UW3/Cx)

N;Alternate names: chlpn 76kda homolog CT622

C;Species: Chlamydia trachomatis

C;Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999

C;Accession: G71490

R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998

A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tract

A;Reference number: A71570; MUID:99000809; PMID:9784136

A;Accession: G71490

A;Molecule type: DNA

A;Residues: 1-647 <ARN>

A;Cross-references: GB:AE001333; GB:AE001273; NID:G3329068; PIDN:AAC68226.1; PID:G332906;

A;Experimental source: serotype D, strain UW-3/Cx

C;Comment: This sequence was originally identified as homologous to part of a sequence (PIR:E72042).

C;Genetics:

A;Gene: CT622

Query Match 29.5%; Score 56; DB 2; Length 647;

Best Local Similarity 33.3%; Pred. No. 71;

Matches 16; Conservative 9; Mismatches 13; Indels 10; Gaps 1;

Qy 1 MDWLKARVEQ-----ELQALEARGTDSNAELRAMEAKLKAEIQ 38

| : : | ||||| | : : | |

Db 364 MQGFRSMIEQFNVPATATAKELQAMEAQLTAMSDQLVGADGELPAEIQ 411

Search completed: January 30, 2004, 13:22:45

Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 13:14:48 ; Search time 41 Seconds
(without alignments)
150.984 Million cell updates/sec

Title: US-10-025-598-2
Perfect score: 190
Sequence: 1 MDWLKARVEQELQALEARGTDSNAELRAMEAKLKAEIQK 39

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 19Jun03:*

1:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	60.5	31.8	794	AAU11437	Mouse hyaluronic a
2	59.5	31.3	49	AAR31987	SSP polypeptide pr
3	59.5	31.3	49	AAR32687	SSP polypeptide pr
4	59.5	31.3	49	AAR32692	SSP polypeptide pr
5	59.5	31.3	49	AAR78249	SSP 7.7.7.7.5.
6	59.5	31.3	49	AAR78239	SSP 5.7.7.7.7.5.
7	59.5	31.3	49	AAR78240	SSP 5.7.7.7.7.5.
8	59.5	31.3	49	AAW60515	Synthetic storage
9	59.5	31.3	49	AAW60505	Synthetic storage

10	59.5	31.3	49	19	AAW60506	Synthetic storage
11	59.5	31.3	49	20	AAW87741	Synthetic lysine-r
12	59.5	31.3	49	20	AAW87742	Synthetic lysine-r
13	59.5	31.3	49	20	AAW87747	Synthetic lysine-r
14	59.5	31.3	56	14	AAR32702	SSP-3-5 polypeptid
15	59.5	31.3	56	14	AAR32685	SSP5-3-3(A/E) poly
16	59.5	31.3	56	16	AAR78260	SSP 5.5.5.7.7.7.7.
17	59.5	31.3	56	19	AAW60526	Synthetic storage
18	59.5	31.3	56	20	AAW87755	Synthetic lysine-r
19	59.5	31.3	77	14	AAR32698	SSP polypeptide pr
20	59.5	31.3	77	16	AAR78258	SSP 7.7.7.7.7.7.8.
21	59.5	31.3	77	19	AAW60524	Synthetic storage
22	59.5	31.3	77	20	AAW87753	Synthetic lysine-r
23	58.5	30.8	224	22	AAW38914	Human polypeptide
24	58.5	30.8	239	21	AAW43319	Human ORFX ORF3083
25	58.5	30.8	254	22	AAW40700	Human polypeptide
26	58.5	30.8	263	21	AAW56890	Human prostate can
27	58.5	30.8	705	23	ABB92587	Herbicidally activ
28	58	30.5	30	14	AAW40051	ACID-pl. Syntheti
29	58	30.5	30	19	AAW54315	pACID leucine zipp
30	58	30.5	30	23	AAE21138	Acidic peptide use
31	58	30.5	31	22	AAE02627	Leucine zipper #1
32	58	30.5	35	23	AAU09852	Antigen presenting
33	58	30.5	39	21	AAW53262	Plasmid pCRII amin
34	58	30.5	39	22	AAW68084	Synthetic peptide
35	58	30.5	39	22	AAW68086	Amino acid sequen
36	58	30.5	45	20	AAW27117	Coiled coil sequen
37	58	30.5	46	21	AAW53264	Acid helper phage
38	58	30.5	47	21	AAW03975	Leucine zipper cas
39	58	30.5	494	22	AAW67480	A fusion protein c
40	57.5	30.3	42	14	AAR32688	SSP polypeptide pr
41	57.5	30.3	42	16	AAR78241	SSP 5.7.7.7.7.5.
42	57.5	30.3	42	19	AAW60507	Synthetic storage
43	57.5	30.3	42	20	AAW87743	Synthetic lysine-r
44	57.5	30.3	44	21	AAW95964	Construct used in
45	57.5	30.3	56	14	AAR32699	SSP polypeptide pr

ALIGNMENTS

RESULT 1

AAU11437	ID	AAU11437	standard; Protein; 794 AA.
XX	XX	AAU11437;	
AC	AAU11437;		
XX	DT	12-MAR-2002	(first entry)
XX	XX	Mouse hyaluronic acid binding protein RHAMM.	
DE	XX	Mouse; hyaluronic acid binding protein; RHAMM; gene therapy; receptor for HA mediated mobility; immunosuppressive; cytostatic.	
KW	KW	conjugate; rheumatoid arthritis; scleroderma; liver fibrosis; cancer;	
XX	OS	Mus musculus.	
XX	XX		
FH	Key	Location/Qualifiers	
FT	Misc-difference	55	/note= "Encoded by AAC"
FT	Misc-difference	71	/note= "Encoded by ACG"
FT	Misc-difference	89	/note= "Encoded by CAA"
FT	Misc-difference	91	/note= "Encoded by CAC"
FT	Misc-difference	540	/note= "Encoded by ACT"
FT	Misc-difference	668	/note= "Encoded by GAT"
XX	WO200180899-A2.		

10/025,598
Search Notes
2/21/04

PT expressed in vivo in plants to serve as custom-tailored storage
PT proteins with specified aminoacid content
XX
PS Disclosure; Page 117; 176pp; English.
XX
CC The sequence represents a synthetic heptad polypeptide which can be
CC expressed in vivo in plants to serve as a synthetic seed storage
CC protein which can be custom-tailored for specific end-user requirements.
CC The DNA encoding the heptad may be used to transform plants to increase
CC the content of partic. amino acids such as lysine or methionine in seeds
CC or leaves. See also AAR31979-86 and AAR32659-705.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 49 AA;

Query Match 31.3%; Score 59.5; DB 14; Length 49;
Best Local Similarity 44.4%; Pred. No. 1.5;
Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

QY 4 LKARVEQELQALEARGTDSNAELRAMEAKLKAEIQK 39
||| :||:|:| : :||| ||| :|
Db 12 LKA-MEEKLKAMEEKLKAMEEKLKAMEEKLKAMEEK 46

RESULT 4
AAR32692
ID AAR32692 standard; Protein; 49 AA.
XX
AC AAR32692;
XX
DT 25-MAR-2003 (updated)
DT 22-JUN-1993 (first entry)
XX
DE SSP polypeptide produced from clone 82-4.
XX
KW Heptad; plants; custom tailored storage proteins; in vivo;
KW expression.
XX
OS Synthetic.
XX
PN WO9303160-A1.
XX
PD 18-FEB-1993.
XX
PF 07-AUG-1992; 92WO-US06412.
XX
PR 09-AUG-1991; 91US-0743006.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Falco SC, Keeler SJ, Rice JA;
XX
DR WPI; 1993-076517/09.
DR N-PSDB; AAQ37273.
XX
PT Synthetic polypeptide(s) contg. specified heptad units -
PT expressed in vivo in plants to serve as custom-tailored storage
PT proteins with specified aminoacid content
XX
PS Disclosure; Page 123; 176pp; English.
XX

CC The sequence represents a synthetic heptad polypeptide which can be
CC expressed in vivo in plants to serve as a synthetic seed storage
CC protein which can be custom-tailored for specific end-user requirements.
CC The DNA encoding the heptad may be used to transform plants to increase
CC the content of partic. amino acids such as lysine or methionine in seeds
CC or leaves. See also AAR31979-86 and AAR32659-705.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 49 AA;

Query Match 31.3%; Score 59.5; DB 14; Length 49;
Best Local Similarity 44.4%; Pred. No. 1.5;

Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

QY 4 LKARVEQELQALEARGTDSNAELRAMEAKLKAEIQK 39
||| :||:|:| : :||| ||| :|
Db 12 LKA-MEEKLKAMEEKLKAMEEKLKAMEEKLKAMEEK 46

RESULT 5
AAR78249
ID AAR78249 standard; Protein; 49 AA.
XX
AC AAR78249;
XX
DT 15-JUL-1996 (first entry)
XX
DE SSP 7.7.7.7.7.7.5.
XX
KW Lysine; synthetic storage protein; SSP; vector; PSK6;
KW dihydrodipicolinic acid synthase; corn; maize; Zea mays;
KW soybean; Glycine max; transgenic plant; essential amino acid.
XX
OS Synthetic.
XX

FH Key Location/Qualifiers
FT Peptide 1..7
FT Peptide /label= SSP7
FT Peptide 8..14
FT Peptide /label= SSP7
FT Peptide 15..21
FT Peptide /label= SSP7
FT Peptide 22..28
FT Peptide /label= SSP7
FT Peptide 28..35
FT Peptide /label= SSP7
FT Peptide 35..42
FT Peptide /label= SSP7
FT Peptide 43..49
FT Peptide /label= SSP5
XX
PN WO9515392-A1.
XX
PD 08-JUN-1995.
XX
PF 21-NOV-1994; 94WO-US13190.
XX
PR 17-JUN-1994; 94US-0261661.
PR 30-NOV-1993; 93US-0160117.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Falco SC, Keeler SJ, Rice JA;
XX
DR WPI; 1995-215272/28.
DR N-PSDB; AAQ94992.
XX
PT New chimeric gene providing increased lysine content in plant seeds
PT - contains di:hydro:di:picolinic acid synthase gene coupled to
PT chloroplast transport sequence and seed specific promoter, also new
PT plants of improved nutritional value.
XX
PS Example 8; Page 135; 180pp; English.
XX

CC Lysine-rich synthetic storage protein SSP 7.7.7.7.7.7.5 (AAR78249)
CC comprises 6 heptad SSP7 repeats (see AAR78237) and a C-terminal SSP5
CC heptad (see AAR78236) derived from vector PSK6; an N-terminal SSP5
CC was lost owing to homologous recombination between base gene
CC repeats. The SSP was obt'd. by insertion of SSP7-encoding
CC oligonucleotides (AAQ94976-77) into PSK6 and use of the resulting vector
CC to transform Escherichia coli JM103, yielding clone 82-4 (AAQ94992).
CC The SSP forms a coiled-coil structure. It can be expressed in the
CC seeds of transformed plants, e.g. soybean and corn, to increase
CC lysine content.
XX

SQ Sequence 49 AA; Query Match 31.3%; Score 59.5; DB 16; Length 49; Best Local Similarity 44.4%; Pred. No. 1.5; Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

QY 4 LKARVEQELQALEARGTDSNAELRAMEAKLKAEIQK 39
||| :||:|:| : :||| |||| :|
Db 5 LKA-MEEKLKAMEEKLKAMEEKLKAMEEKLKAMEEK 39

RESULT 6
AAR78239
ID AAR78239 standard; Protein; 49 AA.
XX
AC AAR78239;
XX
DT 15-JUL-1996 (first entry)
XX
DE SSP 5.7.7.7.7.5.
XX
KW Lysine; synthetic storage protein; SSP; vector; pSK6;
KW dihydrodipicolinic acid synthase; corn; maize; Zea mays;
KW soybean; Glycine max; transgenic plant; essential amino acid.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1..7
FT /label= SSP5
FT Peptide 8..14
FT /label= SSP7
FT Peptide 15..21
FT /label= SSP7
FT Peptide 22..28
FT /label= SSP7
FT Peptide 28..35
FT /label= SSP7
FT Peptide 35..42
FT /label= SSP7
FT Peptide 43..49
FT /label= SSP5
XX
PN WO9515392-A1.
XX
PD 08-JUN-1995.
XX
PF 21-NOV-1994; 94WO-US13190.
XX
PR 17-JUN-1994; 94US-0261661.
PR 30-NOV-1993; 93US-0160117.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Falco SC, Keeler SJ, Rice JA;
XX
DR WPI; 1995-215272/28.
DR N-PSDB; AAQ94978.
XX
PT New chimeric gene providing increased lysine content in plant seeds
PT - contains di:hydro:di:picolinic acid synthase gene coupled to
PT chloroplast transport sequence and seed specific promoter, also new
PT plants of improved nutritional value.
XX
PS Example 8; Page 123; 180pp; English.
XX
CC Lysine-rich synthetic storage protein SSP 5.7.7.7.7.5 (AAR78239)
CC comprises 5 heptad SSP7 repeats (see AAR78237) flanked by SSP5 heptads
CC (see AAR78236) derived from vector pSK6. It was obtd. by insertion of
CC SSP7-encoding oligonucleotides (AAQ94976-77) into pSK6 and use of
CC the resulting vector to transform Escherichia coli JM103, yielding
CC clone C15 (AAQ94978). The SSP forms a coiled-coil structure. It can
CC be expressed in the seeds of transformed plants, e.g. soybean and corn,

CC to increase lysine content.
XX
SQ Sequence 49 AA; Query Match 31.3%; Score 59.5; DB 16; Length 49; Best Local Similarity 44.4%; Pred. No. 1.5; Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

QY 4 LKARVEQELQALEARGTDSNAELRAMEAKLKAEIQK 39
||| :||:|:| : :||| |||| :|
Db 12 LKA-MEEKLKAMEEKLKAMEEKLKAMEEKLKAMEEK 46

RESULT 7
AAR78240
ID AAR78240 standard; Protein; 49 AA.
XX
AC AAR78240;
XX
DT 15-JUL-1996 (first entry)
XX
DE SSP 5.7.7.7.7.5.
XX
KW Lysine; synthetic storage protein; SSP; vector; pSK6;
KW dihydrodipicolinic acid synthase; corn; maize; Zea mays;
KW soybean; Glycine max; transgenic plant; essential amino acid.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1..7
FT /label= SSP5
FT Peptide 8..14
FT /label= SSP7
FT Peptide 15..21
FT /label= SSP7
FT Peptide 22..28
FT /label= SSP7
FT Peptide 29..35
FT /label= SSP7
FT Peptide 36..42
FT /label= SSP7
FT Peptide 43..49
FT /label= SSP5
XX
PN WO9515392-A1.
XX
PD 08-JUN-1995.
XX
PF 21-NOV-1994; 94WO-US13190.
XX
PR 17-JUN-1994; 94US-0261661.
PR 30-NOV-1993; 93US-0160117.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Falco SC, Keeler SJ, Rice JA;
XX
DR WPI; 1995-215272/28.
DR N-PSDB; AAQ94979.
XX
PT New chimeric gene providing increased lysine content in plant seeds
PT - contains di:hydro:di:picolinic acid synthase gene coupled to
PT chloroplast transport sequence and seed specific promoter, also new
PT plants of improved nutritional value.
XX
PS Example 8; Page 124; 180pp; English.
XX
CC Lysine-rich synthetic storage protein SSP 5.7.7.7.7.5 (AAR78240)
CC comprises 5 heptad SSP7 repeats (see AAR78237) flanked by SSP5 heptads
CC (see AAR78236) derived from vector pSK6. It was obtd. by insertion of
CC SSP7-encoding oligonucleotides (AAQ94976-77) into pSK6 and use of
CC the resulting vector to transform Escherichia coli JM103, yielding
CC clone C15 (AAQ94978). The SSP forms a coiled-coil structure. It can
CC be expressed in the seeds of transformed plants, e.g. soybean and corn,

XX 29-MAR-1999 (first entry)
XX Synthetic lysine-rich storage protein 5.7.7.7.7.5.
DE Lysine; transgenic plant; seed storage protein.
XX Synthetic.
XX

Key Location/Qualifiers
Peptide 1..7
/label= SSP5
Peptide 8..14
/label= SSP7
Peptide 15..21
/label= SSP7
Peptide 22..28
/label= SSP7
Peptide 28..35
/label= SSP7
Peptide 36..42
/label= SSP7
Peptide 43..49
/label= SSP5

XX WO9842831-A2.

PN 01-OCT-1998.

XX 27-MAR-1998; 98WO-US06051.

XX 27-MAR-1997; 97US-0824627.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Epelbaum SU, Falco SC, McDevitt RE;

XX WPI; 1999-045139/04.

XX N-PSDB; AAV99514.

XX Nucleic acids and chimeric genes for increasing seed lysine content
PT - comprise sequence encoding all or part of lysine ketoglutarate
PT reductase, useful to improve nutritional quality of seeds from
PT transformed plants

XX Example 21; Page 135; 231pp; English.

XX This is the amino acid sequence of a lysine-rich synthetic seed
CC storage protein (SSP) composed of SSP5 and SSP7 heptad repeats
CC (sequence by heptad repeat is 5.7.7.7.7.5). A synthetic gene
CC (see AAV99514) encoding the SSP was constructed by insertion of
CC SSP7-encoding oligonucleotides (see AAV99509) into a unique EarI
CC site in the SSP5-encoding 'base gene' (see AAV99505) of vector pSK5.
CC Chimeric genes for lysine-rich synthetic seed storage proteins
CC suitable for expression in the seeds of plants have been
CC constructed. The invention provides methods for improving the
CC nutritional quality of seeds from transgenic plants by increasing
CC lysine content.

XX Sequence 49 AA;

Query Match 31.3%; Score 59.5; DB 20; Length 49;
Best Local Similarity 44.4%; Pred. No. 1.5;
Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

Qy 4 LKARVEQELQALEARGTDSNAELRAMEAKLKAEIQK 39

Db 12 LKA-MEEKLKAMEEKLKAMEEKLKAMEEKLKAMEEK 46

RESULT 13

AAW87747

ID AAW87747 standard; Protein; 49 AA.

XX AAW87747;
XX 29-MAR-1999 (first entry)
DT Synthetic lysine-rich storage protein 7.7.7.7.7.5.
XX Lysine; transgenic plant; seed storage protein.
XX Synthetic.
XX

Key Location/Qualifiers
Peptide 1..7
/label= SSP7
Peptide 8..14
/label= SSP7
Peptide 15..21
/label= SSP7
Peptide 22..28
/label= SSP7
Peptide 29..35
/label= SSP7
Peptide 36..42
/label= SSP7
Peptide 43..49
/label= SSP5

XX WO9842831-A2.

PN 01-OCT-1998.

XX 27-MAR-1998; 98WO-US06051.

XX 27-MAR-1997; 97US-0824627.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Epelbaum SU, Falco SC, McDevitt RE;

XX WPI; 1999-045139/04.

XX N-PSDB; AAV99527.
XX Nucleic acids and chimeric genes for increasing seed lysine content
PT - comprise sequence encoding all or part of lysine ketoglutarate
PT reductase, useful to improve nutritional quality of seeds from
PT transformed plants

XX Example 21; Page 144; 231pp; English.

XX This is the amino acid sequence of a lysine-rich synthetic seed
CC storage protein (SSP) composed of SSP7 and SSP5 heptad repeats
CC (sequence by heptad repeat is 7.7.7.7.7.5). A synthetic gene
CC (see AAV99527) encoding the SSP was constructed by insertion of
CC SSP7-encoding oligonucleotides (see AAV99509) into a unique EarI
CC site in the SSP5-encoding 'base gene' (see AAV99505) of vector pSK6.
CC Chimeric genes for lysine-rich synthetic seed storage proteins
CC suitable for expression in the seeds of plants have been
CC constructed. The invention provides methods for improving the
CC nutritional quality of seeds from transgenic plants by increasing
CC lysine content.

XX Sequence 49 AA;

Query Match 31.3%; Score 59.5; DB 20; Length 49;
Best Local Similarity 44.4%; Pred. No. 1.5;
Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

Qy 4 LKARVEQELQALEARGTDSNAELRAMEAKLKAEIQK 39

Db 5 LKA-MEEKLKAMEEKLKAMEEKLKAMEEKLKAMEEK 39

RESULT 14

```
AAR32702
ID AAR32702 standard; peptide; 56 AA.
XX AC AAR32702;
XX DT 25-MAR-2003 (updated)
XX DT 22-JUN-1993 (first entry)
XX DE SSP-3-5 polypeptide.
XX KW Heptad; plants; custom tailored storage proteins; in vivo;
XX KW expression; seed storage protein.
XX OS Synthetic.
XX PN WO9303160-A1.
XX PD 18-FEB-1993.
XX PF 07-AUG-1992; 92WO-US06412.
XX PR 09-AUG-1991; 91US-0743006.
XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX PI Falco SC, Keeler SJ, Rice JA;
XX SQ Sequence 56 AA;
WPI; 1993-076517/09.
DR N-PSDB; AAQ37285.
XX Synthetic polypeptide(s) contg. specified heptad units -
PT expressed in vivo in plants to serve as custom-tailored storage
PT proteins with specified aminoacid content
XX Disclosure; Page 143; 176pp; English.
XX The sequence represents a synthetic heptad polypeptide which can be
CC expressed in vivo in plants to serve as a synthetic seed storage
CC protein which can be custom-tailored for specific end-user requirements.
CC The DNA encoding the heptad may be used to transform plants to increase
CC the content of partic. amino acids such as lysine or methionine in seeds
CC or leaves. See also AAR31979-86 and AAR32659-705.
XX (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 56 AA;
Query Match 31.3%; Score 59.5; DB 14; Length 56;
Best Local Similarity 44.4%; Pred. No. 1.8;
Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;
QY 4 LKARVEQELQALEARGTDSNAELFRAMEAKLKAEIQK 39
Db 5 LKA-MEEKLKAMEEKLKAMEEKLKAMEEK 39
RESULT 15
AAR32685
ID AAR32685 standard; peptide; 56 AA.
XX AC AAR32685;
XX DT 25-MAR-2003 (updated)
XX DT 22-JUN-1993 (first entry)
XX DE SSP5-3-3(A/E) polypeptide.
XX KW Heptad; plants; custom tailored seed storage proteins.
XX OS Synthetic.
XX PN WO9303160-A1.
XX PD 18-FEB-1993.
AAR32702
XX PF 07-AUG-1992; 92WO-US06412.
XX PR 09-AUG-1991; 91US-0743006.
XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX PI Falco SC, Keeler SJ, Rice JA;
XX WPI; 1993-076517/09.
XX Synthetic polypeptide(s) contg. specified heptad units -
PT expressed in vivo in plants to serve as custom-tailored storage
PT proteins with specified aminoacid content
XX Claim 7; Page 154; 176pp; English.
XX The sequence represents a synthetic polypeptide comprising heptad
CC units of the peptide. The synthetic polypeptide can be expressed in
CC vivo in plants to serve as a synthetic seed storage protein which
CC can be custom-tailored for specific end-user requirements. The DNA
CC encoding the heptad may be used to transform plants to increase the
CC content of partic. amino acids such as lysine or methionine in seeds
CC or leaves. See also AAR31979-87 and AAR32659-705.
XX (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 56 AA;
Query Match 31.3%; Score 59.5; DB 14; Length 56;
Best Local Similarity 44.4%; Pred. No. 1.8;
Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;
QY 4 LKARVEQELQALEARGTDSNAELFRAMEAKLKAEIQK 39
Db 5 LKA-MEEKLKAMEEKLKAMEEKLKAMEEK 39
Search completed: January 30, 2004, 13:21:00
Job time : 42 secs
```